Glucose dehydrogenase from the hyperthermophilic archaeon Sulfolobus solfataricus

Heyer, Narinder Isabel

Award date:
1999

Awarding institution:
University of Bath

GLUCOSE DEHYDROGENASE FROM THE HYPERTHERMOPHILIC ARCHAEON *SULFOLOBUS SOLFATARICUS*

Submitted by Narinder Isabel Heyer

for the degree of Ph.D. of the

University of Bath, 1999

Narinder I. Heyer

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Contents

List of Figures

List of Tables

Abbreviations

Glycine Gly G	Alanine Ala A	Valine Val V	Leucine Leu L	Isoleucine Ile I
Serine Ser S	Threonine Thr T	Cysteine Cys C	Methionine Met M	Proline Pro P
Aspartic acid Asp D	Asparagine Asn N	Glutamic acid Glu E	Glutamine Gln Q	Lysine Lys K
Arginine Arg R	Histidine His H	Phenylalanine Phe F	Tyrosine Tyr Y	Tryptophan Trp W

A_{600}	Absorbance at 600 nm
ADH	alcohol dehydrogenase (EC 1.1.1.1)
B. megetarium	*Bacillus megaterium*
B. sterothermophilus	*Bacillus sterothermophilus*
B. subtilius	*Bacillus subtilis*
BLUE	F3GA cibacron blue chromatography
bp	base pairs
BSA	bovine serum albumin
CE	cell extract
Ci	Curie (2.2×10^6 decompositions per second)
CS	citrate synthase (EC 4.1.3.7)
DHLipDH	dihydrolipoamide dehydrogenase (EC 1.8.1.4)
ddATP	dideoxyadenosine triphosphate
ddCTP	dideoxycytosine triphosphate
ddGTP	dideoxyguanosine triphosphate
ddTTP	dideoxythymidine triphosphate
DNA	deoxyribonucleic acid
E.coli	*Escherichia coli*
EDTA	(disodium) ethylenediamine tetraacetate
EPPS	N-(2-Hydroxyethyl)piperazine-N'-(3-propanesulfonic acid)
Fasta3	Programme to scan a protein or DNA sequence library for similar sequences
FPLC	fast protein liquid chromatography

g	Relative centrifugal force
GCG	University of Wisconsin Genetics Computer Group
GDH	glucose dehydrogenase (dinucleotide-dependent – EC 1.1.1.47 and PQQ-dependent - EC 1.1.99.17)
GF	gel filtration chromatography
HCl	hydrochloric acid
HEPES	N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
H.mediterannei	*Haloferax mediterannei*
INFAC	Incomplete factorial
IPTG	isopropyl-β-D-thiogalactoside
kb	kilobase pairs
kD	kilodaltons
k_{CAT}	Catalytic rate constant
K_M	Michaelis constant
LADH	horse liver alcohol dehydrogenase
LDH	lactate dehydrogenase (EC 1.1.1.27)
LB	Luria-Bertani
MDR	Medium-chain alcohol/polyol dehydrogenase/reductase family
MES	2-(N-Morpholino)ethanesulfonic acid
$MgCl_2$	Magnesium chloride
MOPS	3-(N-morpholino) propane-sulphonic acid
Mr	Relative molecular mass (molecular weight)
NAD^+	nicotinamide adenine dinucleotide (oxidised form)
NADH	nicotinamide adenine dinucleotide (reduced form)
$NADP^+$	nicotinamide adenine dinucleotide phosphate (oxidised form)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced form)
NaOH	sodium hydroxide
ND	not determined
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
PHA	polyhydroxyalkonoate
PHB	poly-β-hydroxuybutyrate
PHBV	poly(3-hydroxybutyrate-co-3-hydroxyvalerate)
PHV	polyhydroxyvalerate
PMSF	phenylmethane sulphonyl fluoride
P.putida	*Pseudomonas putida*
PQQ	pyrollo-quinoline quinone

PVDF	polvinylidene fluoride
RG	red gel A chromatography
RNAse A	ribonuclease A (EC 3.1.27.5)
rRNA	ribosomal RNA
SDH	sorbitol dehydrogenase (EC 1.1.1.14)
SDS	sodium dodecyl sulphate
SSC	salt/sodium citrate
S. pombe	*Schizosaccharomyces pombe*
S.solfataricus	*Sulfolobus solfataricus*
SWALL	SWISS-PROT All library
$t_{1/2}$	half-life
T. acidophilum	*Thermoplasma acidophilum*
TAE	tris-acetate EDTA
TEMED	N, N, N', N' tetramethylethylene diamine
T. tenax	*Thermoproteus tenax*
Tris	tris-(hydroxymethyl)-methylamine
V_{MAX}	maximum velocity
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

Acknowledgements

Firstly, I would like to thank and acknowledge Kam for the tremendous amount of support he has given me over the years and for getting me here in the first place, and also for coping so well with the stress during the writing of this work! You're the best!

Secondly, of course I would like to thank my supervisors, Mike and David, for their helpful encouragement and criticism throughout the duration of my Ph.D. It has certainly made me into a better scientist and provided me with a strong basis for my future career in research.

I would also like to thank the following people for their direct assistance in the production of this work. At Bath University: Drs Ursula Gerike and Helen Connaris , Mr Peter O'Flaherty, and also Dr Susan Crennel whose tremendous patience and amazing knowledge of computers and *T acidophilum* GDH, helped me generate the work shown in Chapter 8 and Prof. Robert Eisenthal for his helpful discussions on "yucky" enzyme kinetics and of course for his unforgettable surveys! In addition, I would like to thank all the members of the lab. 1.33 (past and present), especially Cat and Krish, who have helped advise me and have also become very good friends too! At Oak Ridge National Laboratory, Oak Ridge, Tennessee: Dr Hugh O' Neill and Dr Jonathan Woodward (my supervisor), everyone else at ORNL and all the "fellow foreigners" I met out there! Thanks for allowing me to make hydrogen in your lab. (sorry for all the breakages!) and also for being so welcoming it was a wonderful experience (and see you there next year!). Thanks everyone at Bath and ORNL, it was a pleasure working (and not working) with you!

Lastly, but certainly not least, I would like to thank my family, Jasdev and Carol (Dad and Mum), Jinny, Gavin and Lilly, (our cat!), all the other members of the vast Heyer "clan" and also all my friends back home (Hannah, Anna, Noushin and Lou). Thank you for all your support and encouragement, it certainly kept me going!

Dedicated to Mum and Dad and Kam

Abstract

Sulfolobus solfataricus is a sulphur-oxidizing Archaeon that grows optimally at 80°C and pH 3. *S. solfataricus* metabolises glucose via the non-phosphorylated Entner-Doudoroff pathway. Glucose dehydrogenase (GDH) is the first enzyme in this pathway, catalysing the reaction :- Glucose + $NAD(P)^+$ → Gluconate + NAD(P)H + H^+

The gene for the GDH enzyme has been cloned and sequenced (AJ012093) from a λ EMBL *S. solfataricus* genomic library and the enzyme found to belong to the zinc-requiring members of the pyridine-nucleotide-dependent dehydrogenase "superfamily". The gene was overexpressed in *E. coli* and the recombinant enzyme was purified to homogeneity using a heat step and affinity chromatography. This yielded a divalent ion-requiring homotetrameric enzyme of 164 kDa, with a specific activity of 42 U/mg. The characterisation studies have confirmed the similarity of the recombinant protein to the native enzyme. The GDH was also shown to possess pH and temperature optima of 10.5 and 90°C, respectively, and a very broad sugar specificity that requires aldopyranoses with equatorially oriented non-H substituents at C-2, C-3 and C-4. The sugar-oxidation rates are dependent upon the coenzyme present. It also possessed a superior thermostability to the previously cloned *Thermoplasma acidophilum* GDH. Preliminary crystallisation trials have yielded orthorhombic crystals of GDH, but these have not provided high resolution data.

GDH can be employed in an established three-tier system to produce molecular hydrogen from biomass. This had previously been established through the enzymatic hydrolysis of biomass to glucose and its subsequent conversion to gluconate and NADPH by *T. acidophilum* GDH (optimal temperature, 59°C). A third step involved the re-oxidation of NADPH via a hydrogenase (from *P. furiosus*; optimal temperature, 100°C) to generate the molecular hydrogen. This system is economically and environmentally significant in providing a non-polluting fuel from renewable sources. However, in order to increase the efficiency of this conversion and allow the enzymes to function nearer to their optimal temperatures, the GDH from *S. solfataricus* has been substituted for the *T. acidophilum* GDH in the current work.

Substitution of the *S. solfataricus* GDH enzyme for the *T. acidophilum* GDH in the hydrogen production system revealed an overall advantage in the rates and yields of hydrogen produced. Stoichiometric yields of 1 mol of H_2 / mol of glucose were achieved with 10 mM glucose at 70°C using 2 U of GDH and 12 U of hydrogenase. This was achieved with 5 mM glucose at 50°C with *T. acidophilum* GDH using 12 U of GDH and 26 U of hydrogenase. Further analysis of the variables present in the system showed that low initial substrate concentrations gave higher yields and conversion rates of hydrogen with the *S. solfataricus* GDH. However, at higher glucose concentrations, the yield of hydrogen declines so that at an initial glucose

concentration of 50 mM, only 20 % yield of hydrogen was obtained. Raising cofactor concentrations also brought about the same decline in yields. Varying pH and temperature did not affect the overall yields but gave starkly contrasting conversion rates. Studies showed that cofactor instability and enzymic inactivation by gluconic acid production were the main causes of low hydrogen yields.

1 Chapter One

Introduction

1.1 OVERVIEW

The enzyme under investigation is glucose dehydrogenase from the hyperthermophilic Archaeon *Sulpholobus solfataricus*. To begin with, a brief summary of extremophiles and the concept of Archaea is given. This is followed by a description of the determinants of protein thermostability and following that the use of extremophilic enzymes, 'extremozymes', in biocatalysis. Finally, accounts of the structural factors relating the 'Medium-chain dehydrogenase/reductase family are given, with the *a priori* knowledge that glucose dehydrogenase is structurally homologous to the above mentioned family of enzymes.

1.2 EXTREMOPHILES AND THE ORIGIN OF LIFE

1.2.1 EXTREMOPHILES

Micro-organisms called extremophiles thrive in harsh living conditions, for example at temperatures above 100°C in terrestrial hot springs and deep-sea thermal vents, at temperatures below 0°C in arctic waters, at immense pressure under several miles of water on the ocean floor, in the saturated salt environment of the Dead Sea or the Great Salt Lake, at pH values less than 2 in deposits and geothermal sulphur-rich springs, at pH values greater than 11 in sewage sludge, and in highly polluted areas subjected to ionising radiation and high heavy metal concentrations. Extremophiles can generally be divided into groups based on their phenotypes as described in Table 1.1. Extremophilic micro-organisms hail mainly from the domain of Archaea, although several thermophilic genera of eubacteria exist as well as a number of psychrophilic representatives. Also eukaryotes possess a number of examples of extremophilic single- and multi-cellular organisms capable of growth in acidic (Ingledew, 1982) and low temperature environments (Gerday, *et al.*, 1997).

Phenotype	Environment	Typical genera
Thermophilic	55 - 80°C	*Methanobacterium, Thermoplasma, Thermus**, some *Bacillus** species
Hyperthermo -philic	80 - 113°C	*Aquifex**, *Archaeoglobus, Hydrogenbacter**, *Methanothermus, Pyrococcus, Pyrodictum, Pyrolobus, Sulfolobus, Thermococcus, Thermoproteus, Thermotoga**
Psychrophilic	-2 - 20°C	*Alteromonas**, *Psychrobacter**
Barophiles	up to 50 Mpa	*Pyrococcus*
Methanogens	strict anaerobes	*Methanobacterium, Methanococcus, Methanothermus*
Halophilic	2 - 5 M NaCl	*Haloarcula, Halobacterium, Haloferax, Halorubrum*
Acidophilic	pH <4	*Acidianus, Desulfurolobus, Sulfolobus, Thermoplasma, Thiobacillus**
Alkaliphilic	pH >9	*Natronobacterium, Natronococcus,* some *Bacillus** species

Table 1.1 Extremophiles and their environments.

*** indicate the bacterial species, the remaining species all reside in the archaeal domain.**

Adapted from Hough and Danson, (1999).

1.2.2 THREE DOMAINS OF LIFE

Former classification of living things by Linnaeus in the 18th century divided organisms into two kingdoms, animals and plants. In the 19th century Haeckel expanded the classification to three kingdoms to include single-celled protists, and Copeland (1938) later split out a fourth branch in order to accommodate a new kingdom, the bacteria. Whittaker, (1969) then created a fifth, for the fungi. However, these definitions focused only on the morphological diversity of the organisms and neglected the molecular and biochemical diversity now apparent in the eukaryotic and prokaryotic microbes. However, according to Zuckerkandl and Pauling (1965), macromolecular sequence comparisons are the most accurate and reliable basis from which to infer phylogenetic relationships, as they allow straightforward, quantitative interpretation and subsequent referencing to the ever growing database. Therefore, the present perspective on evolutionary history of life was provided by sequence comparisons of small subunit rRNA. This was because it is thought to be one of the "best" molecules for comparative sequence studies in that it does not undergo genetic transfer between species, it has a strictly conserved function and also that it contains a sufficient number of residues, which change at a rate commensurate with its evolutionary distance. Comparison of these sequences proposes a tripartite division of the living world into the following domains, Eukarya, Bacteria and Archaea (Woese and Fox, 1977; Pace *et al.*, 1986; Woese *et al.*, 1990). This finding was incompatible with the eocyte

hypothesis proposed by Lake (1988) and Lake *et al.* (1984), which distributes the various archaebacteria among three separate kingdoms: the 'eocytes' (sulfur-dependent archaebacteria, 'photocytes' (extreme halophiles and eubacteria) and methanogens. However, further evidence of the tripartite grouping was found via rooting of paralogous gene families whose duplication is thought to predate the last common ancestor (Iwabe *et al.*, 1989; Gogarten *et al.*, 1989). This information placed the root of the universal tree in the bacteria (Brown and Doolittle, 1995) and determined archaea and eukaryotes to be sister groups (Fig. 1.1). In fact among the ribosomal proteins there were also cases where Archaeal and eukaryotic homologues have no apparent counterpart among the Eubacteria (Kimura *et al.*, 1989; Auer *et al.*, 1898). However, this consensus has recently been challenged by Gupta and Golding, (1986), who showed that depending on the protein dataset used, the molecular phylogenies obtained show various relationships between Archaea, Eubacteria and Eukaryotes. Archaea appear to share homologies with eukaryotes in their informational genes (those involved in transcription, translation and related processes) and with Eubacteria in their operational genes (those involved in housekeeping). This has led to a number of hypotheses in order to reconcile these differences, which involve fusion, engulfment and symbiosis to produce a chimeric organism (Zillig, 1991; Lake and Rivera, 1994; Martin and Müller, 1998; Doolittle, 1998; Koonin *et al.*, 1997; Gupta and Golding, 1996; Sogin, 1991; Lopez-Garcia and Moreria, 1999), as well as the real possibility of massive horizontal interdomain gene transfer (Maynard *et al.*, 1993; Dröge *et al.*, 1998; Lake *et al.*, 1999; Aravind *et al.*, 1999).

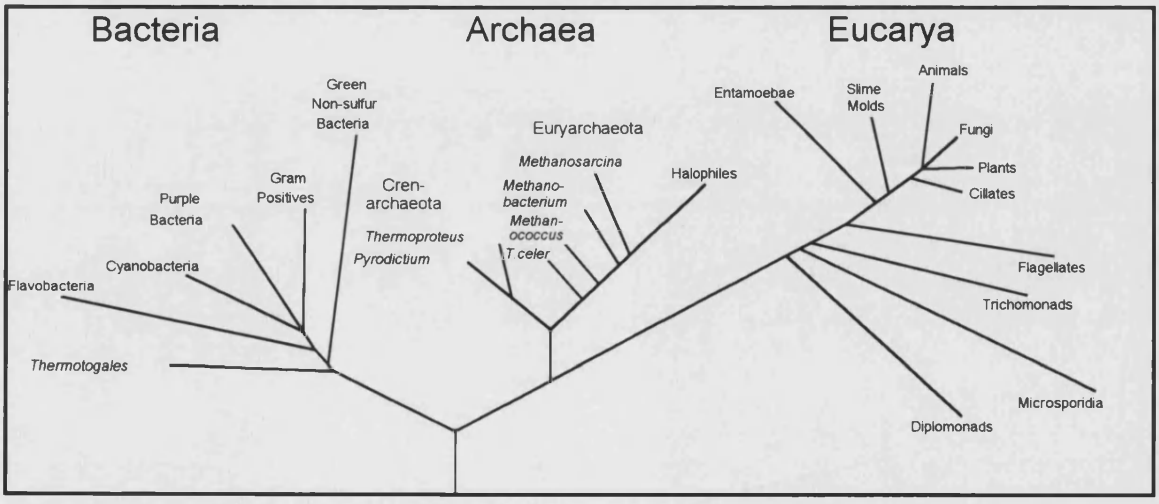


Fig. 1.1 The universal phylogenetic tree, constructed from rRNA sequence comparisons, according to Woese *et al.* (1990).

1.2.3 ARCHAEA

Despite the high degree of homology of archaeal genes and their informational and operational counterparts in eukaryotes and eubacteria, respectively, archaea possess many unique biochemical features that reflect their distinct evolutionary course with respect to either eubacterial or eukaryotic cells. A major difference is displayed by archaeal lipids, which are isopranyl ether-liked glycerolipids, unlike the straight-chain fatty acyl ester-liked glycerolipids present in eubacterial and eukaryotic membranes (Langworthy, 1985). In addition, thermoacidophilic membranes are further modified by the presence of tetraethers (also present in some methanogens) and cyclopentyl rings on the biphytanyl chains (Langworthy, 1985). Archaea are further characterised by their unique cell-envelopes (Kandler and Konig, 1985) and transfer RNA (tRNA) (Gupta, 1985).

Phlogenetically Archaea fall into two major lineages, the kingdoms of crenarchaeota and eukaryarchaeota although a third kingdom, as yet unsubstantiated has been proposed by Barns *et al.*, (1996) (Fig. 1.2). Eukaryarchaeota comprises thermophilic and mesophilic methanogens and sulphur metabolizers, as well as halophiles. Crenarchaeota did comprise exclusively what are known as sulfur-dependent hyperthermophiles or eocytes, although recent molecular phylogenetic surveys of marine planktonic micro-organisms show that in fact crenarchaeota appear to be the most widely distributed, abundant and ecologically diverse of all known archaea (DeLong, 1998a; DeLong, 1998b; Barns *et al.*, 1996). However, these members and also all living organisms are thought to tentatively share a common thermophilic ancestor (DeLong 1998a; Barns *et al.*, 1996).

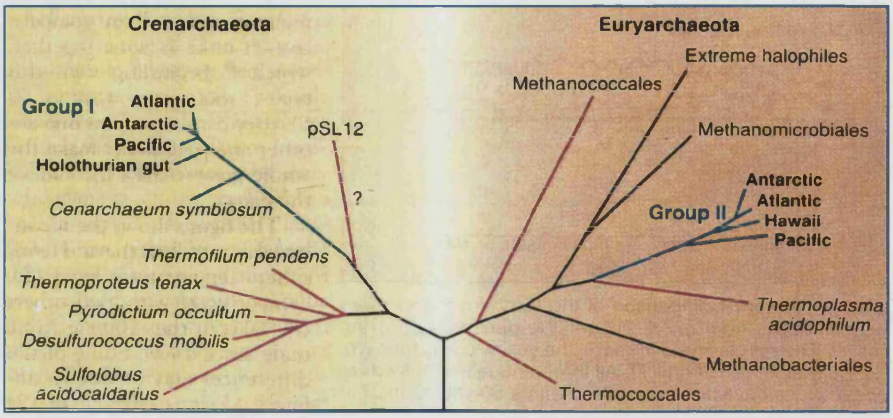


Fig. 1.2 Phylogenetic tree showing Crenarchaeota and Eukaryarchaeota. Reproduced from DeLong, 1998.

Blue lines, cold-dwelling archaea; red lines, predominantly thermophilic groups of cultivated archaea. pSL12 is an rRNA sequence recovered from a presumed thermophile at a Yellowstone hot spring; it is suggested that this organism may belong to the new kingdom of "Korarchaeota" (Barns *et al.* 1996).

1.2.4 THERMOPHILIC ARCHAEA

The hot conditions thought to prevail when life first arose on earth and the deep rooting of thermophilic organisms such as *Aquifex* and *Pyroccocus* suggest that life shares a thermophilic universal ancestor.

The distinct grouping of crenarchaeotic hyperthermophiles does not include the moderate thermophiles *Thermoplasma* and *Thermococcus*, despite their common phenotypes, as 16S rRNA sequence information reveals stronger links with the methanogens and extreme halophiles (Zillig *et al.*, 1985).

Sulfolobus is a sulfur-oxidising genus of archaea that was isolated by Brock *et al.* (1972) from Yellowstone National Park isolates. Additional strains have also been isolated from a number of geographically distant solfatara fields (De Rosa *et al.*, 1974; De Rosa *et al.*, 1975; Segerer and Stetter, 1991; Zillig *et al.*, 1994). *Sulfolobus* thrives in a variety of naturally acidic thermal habitats, both terrestrial and aquatic. It shows optimal growth rates at 75° - 85°C, pH 3 - 4 (Brock *et al.*, 1972). The species used in this work, *Sulfolobus solfataricus*, was previously named *Caldariella acidophila* (De Rosa *et al.*, 1974) and it is a facultative heterotroph capable of

of growing on various complex organic substrates as well as simple sugars and amino acids, and in most cases elemental sulfur (Brock *et al.*, 1972; Segerer and Stetter, 1991).

The moderately thermophilic genus *Thermoplasma*, which is also heavily cited in this work, is also found to inhabit solfataric springs (Segerer *et al.*, 1988), as well as coal refuse sites (Brock, 1986; Segerer *et al.*, 1988). It differs from *Sulfolobus* not only in its phylogenetic grouping but also through its lack of a cell wall and envelope, its lower temperature range (37° - 55°C), its strict heterotrophicity and its inability to oxidise elemental sulfur (Brock, 1972; Segerer and Stetter, 1991). However, in spite of these differences *Sulfolobus* and *Thermoplasma* can be exclusively grouped by their virtually unique mode of hexose metabolism.

Central metabolism can be considered in three different stages, namely the catabolism of hexose sugars to pyruvate, the conversion of pyruvate to acetyl-CoA and the oxidation of acetyl-CoA via the citric acid cycle. The catabolism of hexose sugars to pyruvate can occur via several different pathways depending on the organism and the environmental conditions. In eukaryotes and many anaerobic eubacteria glucose catabolism occurs via the Embden-Meyerhof (EM) gylcolytic pathway. However, in many strictly aerobic eubacteria, glucose is catabolised by the Entner-Doudoroff (ED) pathway (Entner & Doudoroff, 1951), due to the absence of the enzyme 6-phosphofructokinase. Eukaryotes and eubacteria also possess an additional pathway for glucose metabolism, the Pentose phosphate pathway (see Fig. 1.3). However, the initial failure to detect any ATP-phosphofructokinase in any archaeal species spurred the discovery of the modified ED pathways. Extreme halophiles were found to employ a semi-phosphorylated ED pathway (Hochstein, 1988) and the thermophilic archaea *Sulfolobus*, *Pyrococcus*, *Thermoplasma* and *Thermoproteus* were found to use a further modification of this pathway, in which few phosphorylated intermediates are used (Fig. 1.4) (De Rosa *et al.*, 1984; Schäfer and Schönheit, 1992; Budgen and Danson, 1989; Siebers *et al.*, 1997). This non-phosphorylated ED pathway yields no net ATP and energy is therefore thought to be derived from the oxidative citric acid cycle or by oxidation of reduced pyridine nucleotide by oxygen in the respiratory chain (Danson and Hough, 1992; Schäfer *et al.*, 1994). However, further analysis by Selig *et al.* (1997) revealed that of a variety of hyperthermophiles, and particularly the anaerobic archaea, were able to degrade sugar to pyruvate using modified EM and ED

pathways or by a combination of both. The primitive lineages of the organisms studied by Selig *et al.* (1997) have led to more speculation on the origin of hexose catabolism that would have formed the basis for the survival and successful competition of early life on earth.

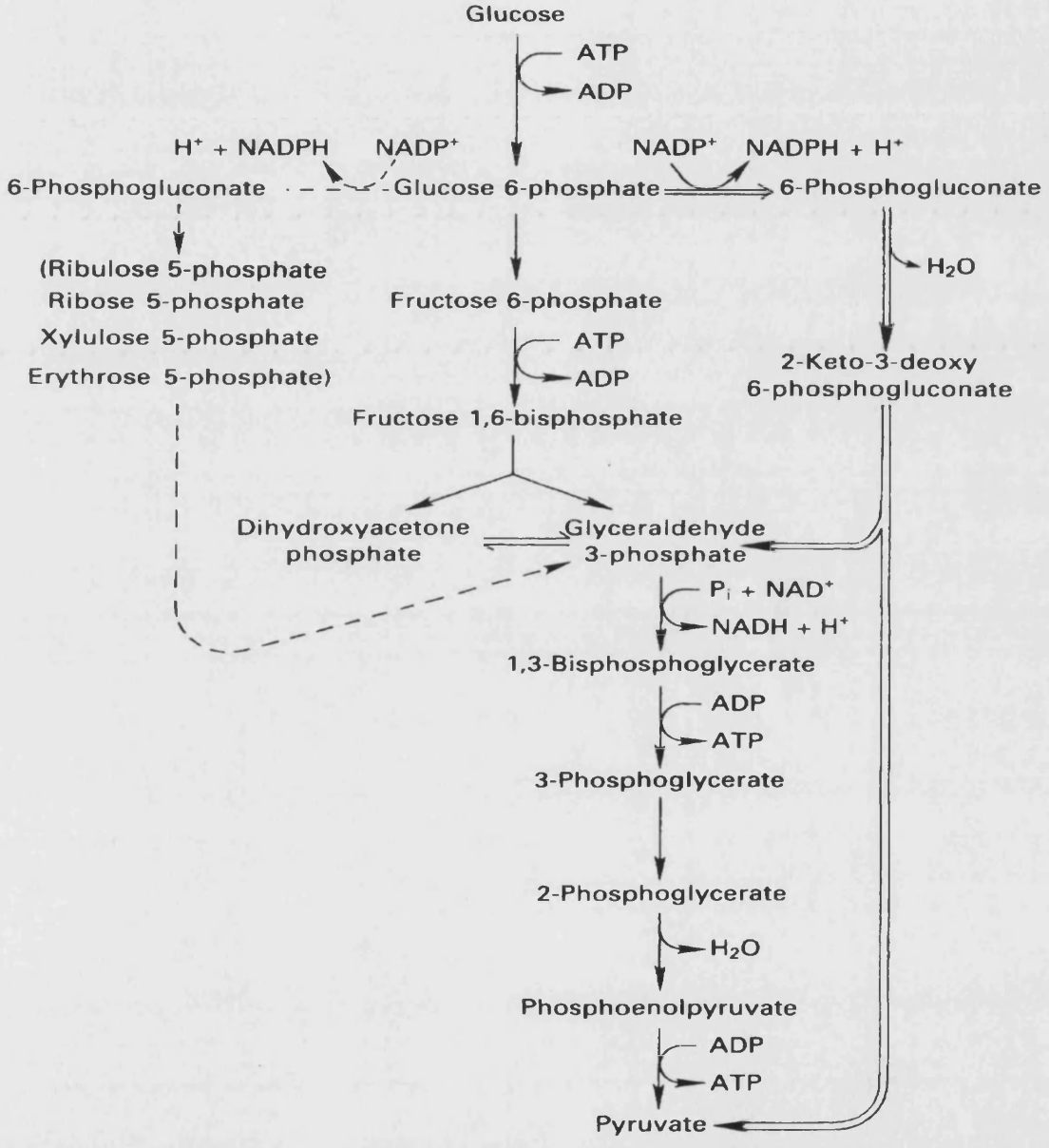


Fig. 1.3 **Pathways of glucose catabolism in eubacteria and eukaryotes. Reproduced from Danson and Hough, 1992.**

The Embden-Meyerhof glycolytic pathway (→), the Entner-Doudoroff pathway (⇒) and the pentose phosphate pathway (- - →).

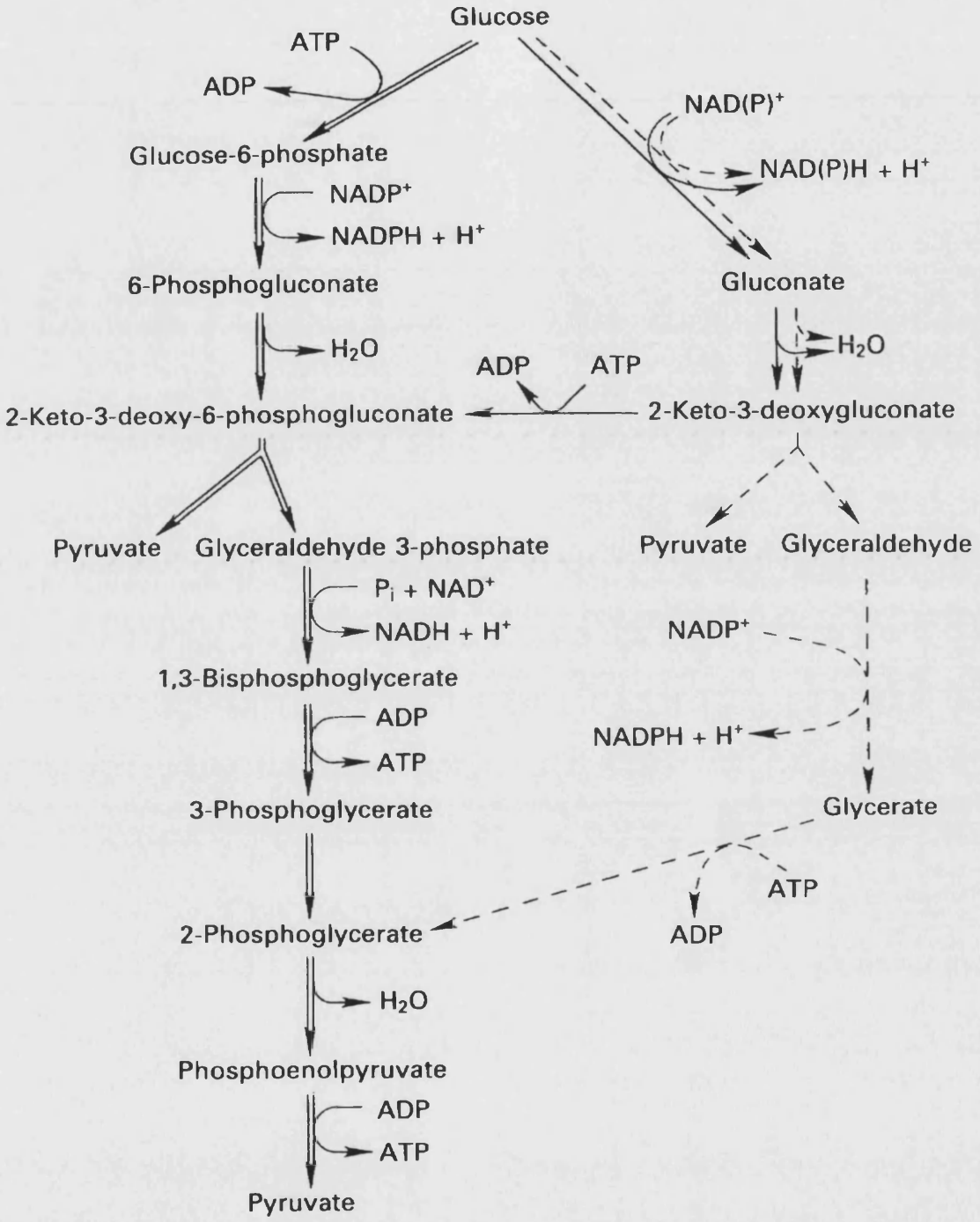


Fig. 1.4 **Pathways of glucose catabolism in halophilic and thermophilic archaebacteria. Reproduced from Danson and Hough, 1992.**

The modified Entner-Doudoroff pathway in halophiles (→) and the non-phosphorylated Entner-Doudoroff pathway in *Sulfolobus solfataricus* and *Thermoplasma acidophilum* (- - →) are shown in comparison with the classical Entner-Doudoroff pathway (⇒).

1.3 ENZYME THERMOSTABILITY

1.3.1 DETERMINING ENZYME THERMOSTABILITY

The discovery and characterisation of hyperthermophilic bacteria capable of growing at temperatures as high as 113°C (Blochl *et al.*, 1997) has led to the understanding that their proteins (especially those that are secreted and therefore unprotected by intracellular solutes (Bragger *et al.*, 1989)) are inherently thermostable and active at the organisms' maximum growth temperature. This makes them very useful tools for studying the structural basis of protein stability, in which two basic approaches have been employed. The first is site directed mutagenesis followed by selection and characterisation and the second is comparative analysis of the three-dimensional structure of homologous proteins from mesophilic and thermophilic sources.

With the exclusion of phylogenetic variations, thermostable enzymes are found to be very similar to their mesophilic counterparts in terms of amino acid sequence (Vielle *et al.*, 1995; Burdette *et al.*, 1996) and three-dimensional structures and catalytic mechanisms (Fujinaga *et al.*, 1993; Russell *et al.*, 1997). It is also possible (with few exceptions) to successfully express these proteins in mesophilic organisms and maintain all their native biochemical properties, thus proving their intrinsic thermostabilty. Therefore, increases in stability seem to come from numerous and subtle interactions, including amino acid exchanges, hydrogen bonds, ionic bonds, hydrophobic interactions, internal packing, stabilisation of helices, surface loop reduction and strategic placement of prolines in β-turns (Jaenicke, 1991; Jaenicke *et al.*, 1990; Matthews, 1993; Vielle and Zeikus, 1996).

1.3.2 STRUCTURAL DETERMINANTS OF THERMOSTABILITY

1.3.2.1 Flexibility vs. rigidity

In addition to enhanced stability at elevated temperatures, thermophilic enzymes also possess a greater resistance to chemical denaturation than their mesophilic counterparts (Colombo *et al.*, 1992; Consalvi *et al.*, 1993; Lebbink *et al.*, 1995). The observed structural resistance may reflect a restriction on the flexibility of thermophilic proteins, presumably through their increased compactness (Russell *et al.*, 1994), which renders them unusually rigid at mesophilic temperatures. However, at their physiological temperatures, the structural rigidity of these proteins is decreased owing to thermal motions and this increased flexibility is

essential to protein function, which invariably requires rapid, reversible changes in protein conformation (Creighton, 1984). However, the degree of flexibility achieved by thermophilic proteins is thought to be less than that attainable by their mesophilic counterparts. Therefore, while allowing them to be functionally competent at elevated temperatures, their catalytic efficiency is significantly compromised. This can be clearly shown when considering the temperature coefficients for enzymic reactions, where the factor by which the rate of a reaction is increased on raising the temperature by 10°C, is usually about two-fold. Therefore, one would predict a more than a sixty-fold higher rate for the thermophilic enzyme at 100°C, than that of the mesophile at 37°C. However, when observing the specific activities of these enzymes at their respective temperature optimums, they are actually found to be of the same order (O'Fágáin, 1995). Therefore, the adaptation of proteins to extreme temperature appears to be the result of a compromise between rigidity responsible for thermal stability and the flexibility required for playing their physiological roles.

1.3.2.2 Amino acid exchanges

Sequence comparison of several thermophilic and mesophilic enzyme homologues has revealed many significant substitutions in the thermophilic enzymes, such as Lys to Arg, Ser to Ala, Gly to Ala, Ser to Thr and Val to Ile (Argos, 1979). Also notable is the reduction and/or increased burial of thermolabile residues (Danson & Hough, 1998).

Comparative studies have also revealed the strategic placing of exchanged amino acids. These exchanges occur by the inclusion of prolines in β-turns (Tanner *et al.*, 1996). Other exchanges occur to increase the internal and decrease the external hydrophobicity and therefore stabilise loops and helices, (Menendez-Arias & Argos, 1988; Querol *et al.*, 1996). However, another area of increased amino acid exchange on increasing thermostability are those that promote an increased number of hydrogen bonds and salt-links (Vogt *et al.*, 1997; Yip *et al.*, 1998).

1.3.2.3 Electrostatic interactions: hydrogen bonds and ion pairs

It was proposed by Perutz & Raidt (1975) and Perutz, (1978) that ion-pairs are an important factor in protein stabilisation. Subsequent work carried out by Vogt *et al.* (1997) showed a significant statistical consistency in the use of hydrogen bonds and salt bridges to increase thermal stability in protein structures taken from sixteen protein families.

Structural comparative work carried out on homologues of two protein families from sources covering a wide range of temperatures (Russell *et al.*, 1994; Russell *et al.*, 1997; Yip *et al.*, 1995) revealed an increase in the incidence of inter-subunit ion-pair networks, in the hyperthermophilic enzymes from *Pyrococcus furiosus.* Ion-pair networks arise from interactions in the assembled enzyme of a patch on the subunit surface that contains a high proportion of charged residues. The ion-pair networks can be quite extensive with one example from the glutamate dehydrogenase from *P. furiosus* involving the interaction of 18 residues across four monomers (Yip *et al.*, 1998). Therefore, Xiao and Honig, (1999), have proposed this to be a key determinant in the stabilisation of proteins from hyperthermophiles.

1.3.2.4 Surface to volume ratio

The surface to volume ratio, expressed in $Å^{-1}$, is the ratio of the area of the molecular surface to the volume enclosed by the molecular surface. This parameter is significantly lowered in thermophilic proteins (Tanner *et al.*, 1996). Thermophilic proteins possess increases in the non-polar and charged surface areas and a decrease in the solvent accessible area (Yip *et al.*, 1995). Furthermore, the polar buried surface areas are often increased in thermophilic proteins (Lim *et al.*, 1997). This parameter also encompasses the increased number of intersubunit ion-pairs and hydrogen bonds, which work to strengthen the intersubunit associations.

1.3.2.5 Hydrophobic interactions

Hydrophobic interactions are considered to be the main driving forces in protein folding (Dill, 1990). In thermophilic proteins the buried nonpolar surface areas are found to be larger than in mesophilic proteins (Lim *et al.*, 1997). This observation, together with work carried out by Pace (1992) to bring about a 1.3 (±0.5) kcal/mol gain in the stability of a four proteins on adding CH_2-groups to aliphatic side-chains, and that carried out by Kirino *et al.* (1994), which determined the increase in hydrophobic intersubunit contacts in thermophilic proteins, indicate the importance of hydrophobic interactions in maintaining the native state.

1.3.2.6 Packing density and compactness

The packing density is defined as the ratio of the volume enclosed by the Van der Waals envelope of a given atom to the actual volume of the space that it occupies. The relatively small solvent-exposed surface area to volume ratio was found to increase the stability

of the aldehyde ferredoxin oxidoreductase enzyme from the hyperthermophile *P. furiosus* by simultaneously reducing the unfavourable surface energy and increasing the attractive interior packing interactions (Chan *et al.*, 1995). Other hyperthermophilic proteins also exhibit increased packing density through the formation of isoleucine clusters (Russell *et al.*, 1997; Knapp *et al.*, 1995). The benefit of employing Ile in favour of other residues can be seen by its mutation to valine, which resulted in a loss of favourable packing interactions of the side-chain in the folded form of a protein (Sneddon and Tobias, 1992). Isoleucine's ability to exist in all three types of χ_1 angles compared to predominantly two for leucine side-chains, thus enables tighter packing and the increased incidence of Van der Waals interactions.

The absence of large cavities in hyperthermophilic proteins is also thought to contribute to their compactness and consequently their degree of hyperthermostability by eliminating areas in the structure where nonoptimal packing occurs (Russell *et al.*, 1997; Delboni *et al.*, 1995; Shin *et al.*, 1996). However, in the case of the structure of β-glycosidase from *S. solfataricus*, a large number of solvent filled hydrophillic cavities were found to exist (Aguillar *et al.*, 1997). It was therefore proposed that this enzyme has employed another method to combat thermal denaturation by providing a degree of "resilience" as opposed to increased rigidity.

The degree of compactness also observed on increasing thermal stability, is furthered by a reduction in surface loops (as highlighted by their increased resistance to proteolysis (Jaenicke, 1991)) and greater complementarity at the subunit interface (Russell *et al., 1997).*

1.3.3 SELECTED EXAMPLES

Many examples of comparative structural data of protein homologues from a series of organisms spanning the entire range of growth temperatures now exist, for example the protein families of citrate synthase, glyceraldehyde-3-phosphate dehydrogenase, superoxide dismutase, adenylate kinase, glutamate dehydrogenase, malate dehydrogenase and DNA polymerase. In order to illustrate how thermostability has been acquired by thermophilic proteins, citrate synthase and glutamate dehydrogenase are briefly presented as examples.

1.3.3.1 Citrate synthase

As shown by the range of citrate synthases (CS) studied and exhibited in Fig. 1.5, the basic structures do not deviate in their overall fold across the variation in physiological temperatures (37° - 100°C). However, on increasing thermostability the following changes were

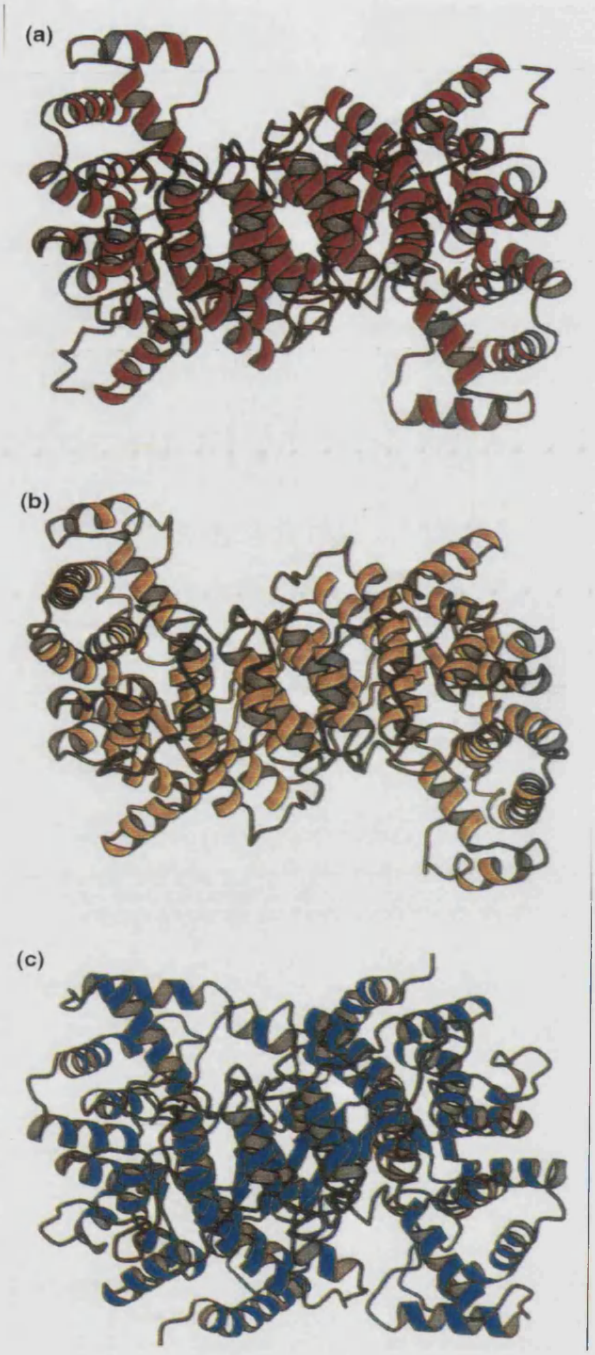


Fig. 1.5 Citrate synthase structures. Schematic representation of the dimeric citrate synthases from (a) *Pyrococcus furiosus* (100°C) (b) *Thermoplasma acidophilum* (55°C) and (c) pig (37°C). Each view is down the two-fold axis and has been produced as described in Russell *et al.* (1998).

surface loop lengths, reduction in internal cavities, increased internal packing (via isoleucine

clusters) and greater complementarity at the subunit interface. The adaptation at the intersubunit contact regions is thought to be particularly significant, with the moderately thermophilic CS possessing more hydrophobic contacts but the hyperthermophilic CS apparently strengthening its subunit interactions by greater numbers of intersubunit ion-pairs and ion-pair networks and an extended carboxyl terminus interaction.

1.3.3.2 Glutamate dehydrogenase

This comparative study employs the crystal structures of glutamate dehydrogenase (GluDH) from the mesophile *Clostridium symbiosum* and from the hyperthermophiles *Pyrococcus furiosus* (growth temperature = 100°C) and *Thermotoga maritima* (growth temperature = 80°C) (Knapp, 1997). The most significant differences on increasing thermostablity were increases ion-pairs and intrasubunit ion-pairs, as well as the decrease in the volume of intersubunit cavities. However, this study reveals interesting differences in the type of subunits interactions employed by the hyperthermophiles, where *P. furiosus* GluDH possessed a substantial number of salt-bridge networks, whereas the *T. maritima* GluDH possessed more hydrophobic interactions between its subunits, and this is precisley what is observed with CS.

1.3.4 EXTRINSIC FACTORS

The ability of the majority of purified proteins to refold spontaneously *in vitro* after being completely unfolded must be dictated by the primary sequence. Therefore, attainment of the three-dimensional structure of a protein is an intrinsic property of the protein itself (Anfinsen, 1973; Creighton, 1990). This conclusion is supported by the ability of genes encoding thermophilic or hyperthermophilic enzymes to be actively expressed in a mesophilic host, such that the recombinant proteins usually have virtually identical properties to those of the native enzyme. However, exceptions to this rule are seen in cases where differences arise through the absence/presence of post-translational modifications or for example, the recombinant aspartate transcarbamylase enzyme from *Pyrococcus abyssi*. This enzyme was found to be less thermo- and pressure stable than its native counterpart and also possessed a lower cooperativity in ligand binding (Purcarea *et al.*, 1997). Yet even in their native habitat, the proteins of the thermophilic archaeal organism *Methanothermus fervidus* are not found to be inherently stable and instead are thought to acquire thermoprotection from high intracellular concentrations of

the anion 2-3-diphosphogylcerate (Hensel and Konig, 1988). Indeed more evidence on the accumulation of extrinsic factors in the form of organic solutes by hyperthermophilic archaea is appearing (Hensel, 1993; Ciulla *et al.*, 1994; Martins *et al.*, 1997) and it is thought that these compounds may have a direct effect on enzyme thermostability. Other low-molecular mass compounds such as ions, metabolites, coenzymes or specific protectants (e.g. glycerol) and stress proteins (Trent *et al.*, 1991) have also been found to act as structure-stabilising factors. Use of these factors may further increase enzyme stability *in vitro*, as well as suggesting another way in which proteins have tried to survive in their extremophilic environments.

1.4 INDUSTRIAL APPLICATION OF ENZYMES

1.4.1 BIOCATALYSIS AND THE POTENTIAL FOR EXTREMOZYMES

Biocatalysis in the food industry has been in place for thousands of years and predates any understanding of the chemistry and biology of the process, for example in the production of beer and cheese. The first stages of industrial biocatalysis and biotransformation were developed in the 19th century, in which the large-scale manufacture of ethanol and ascorbic acid are examples. Biocatalysis is now widely used throughout industry and employing industrial enzymes over chemical catalysts is attractive due to their efficiency and selectivity. This results in relatively pure compounds that are enantiospecific, with fewer unwanted by-products. Enzymes can also minimise production time and waste generation as they bio-degrade after performing their task and they do not require additional compounds and solvents to protect and de-protect specific chemical groups. However, further application and improvement of these processes has been hampered by their volumetric productivity, stability and availability. Therefore, the discovery of micro-organisms that could exist at extremes of temperature, pH, pressure and in low water (high salt) environments represented vast stores of potential biocatalysts, due to their ability to function in the harsh conditions usually employed in industrial processes (Table. 1.2). The use of these enzymes could replace existing processes either containing mesophilic homologues or chemical catalysts, thus improving their efficiency. However, the uniqueness of the thermophilic and hyperthermophilic archaeal micro-organisms that possess novel central metabolic (Danson and Hough, 1992) and lipid biosynthetic pathways (Langworthy, 1985) may lead to the large-scale synthesis of completely unique compounds of major industrial importance.

that possess novel central metabolic (Danson and Hough, 1992) and lipid biosynthetic pathways (Langworthy, 1985) may lead to the large-scale synthesis of completely unique compounds of major industrial importance.

Extremophile group	Extremophile molecules	Applications
Barophiles	Proteases, amylases, cellulases, lipases	Enzymes for proteolysis, food processing, and oil well stimulation
Halophiles	Carotene, glycerol, solutes, membrane constituents	Enzymes for peptide synthesis, waste treatment, and oil recovery; surfactants for pharmaceuticals
Thermo- and hyperthermophiles	Amylases, xylanases, DNA polymerases, proteases	Enzymes for molecular biology (e.g., PCR), biosensors, bioremediation, paper and textile processing
Psychrophiles	Proteases, amylases, lipases, dehydrogenases	Enzymes for laundry detergents, fatty acid synthesis, biosensors, pharmaceutical formulation components
Acidophiles	Sulfur oxidation	Coal desulfurication
Alkaliphiles	Cellulases, proteases, amylases, lipases, cyclodextrins, antibiotics	Enzymes for industrial processes, molecular rearrangements

Table 1.2 Applications of extremozymes. Reproduced from Persidis (1998).

1.4.2 ENZYME PRODUCTION

1.4.2.1 What makes a good biocatalyst ?

The factors outlined in Table 1.3 are the requirements to create a good industrial biocatalyst. These determine whether an enzyme would be cost effective and marketable. Cowan (1992) proposed that rather than replace existing commercial enzymes, which may require costly alteration of equipment, extremozymes would benefit in "niche" applications, where their unique properties would offer significant advantages.

Discovery scientists	Novel science
↓	
Development	Reliable, reproducible and easy process integration
↓	
Production	Minimal commissioning problems, speed to first production
↓	
Company	Good returns on investments, low capital costs
↓	
Consumer	New or better quality products and services at more affordable prices
↓	
Society	Job creation, export creation
↓	
The environment	Use of renewable resources, environmental benefits

Table 1.3 Requirements to create a successful biocatalyst. Reproduced from Cheetam (1998).

However, the feasibility of large-scale biocatalysis using extremozymes requires their availability in pure and sufficient quantities. Two methods are currently available for production of extremozymes. One is the expression of the enzyme's gene in the native extremophilic host, such as the system developed by Jolley *et al.* (1996) to overexpress dihydrolipoamide dehydrogenase (DHLipDH) in *Haloferax volcanni*. However, there is currently a lack of suitable gene transfer systems for many of these extremophilic organisms and, furthermore, the conditions required for the growth of these organisms (see Table 1.1) are incompatible with standard industrial fermentation and downstream processing plant. Therefore, considerable interest has grown in heterologous gene expression using a mesophilic host such as *Escherichia coli*. Such systems have been successfully employed in the expression of many thermophilic enzymes e.g. glucose dehydrogenase from *Thermoplasma acidophilum* (Bright *et al.*, 1993) and a few psychrophiles (at sub-mesophilic temperatures) e.g. citrate synthase from an Antarctic bacterium (Gerike *et al.*, 1997). However, heterologous expression of halophile genes in *E. coli* has not been as straight forward. This has resulted in either a soluble inactive product that required the addition of salt for activity (Bischoff and Rodwell, 1996; Cendrin *et al.*, 1993) or in inclusion bodies that required solubilization and refolding (Blecher *et al.*, 1993; Connaris *et al.*, 1999).

Initial purification steps using the heterologous expression could exploit the mesophilic nature of *E. coli* by subjecting it to thermophilic and halophilic conditions in order to denature the host mesophilic proteins and solubilize the respective extremozymes. However, subsequent purification steps of halophilic enzymes are complicated by their requirement for high salt

concentrations for stability, although they have successfully been purified using hydrophobic interaction and affinity chromatography (Mevarech *et al.*, 1976; Sundquist and Fahey, 1988). Thermophilic and psychrophilic enzymes are able to employ conventional purification methods without the need for modification.

1.4.3 SCREENING FOR NOVEL ENZYMES

Given that extremophiles are, by definition, difficult to culture in the laboratory and for the same reasons <1% of the microbial diversity of the planet has been explored (Hugendholtz and Pace, 1996), the key challenge to industry is to isolate and characterise potentially useful molecules from such organisms.

1.4.3.1 Isolating specific enzymes

If a specific enzyme from a well-known and characterised protein class is required, then gene probes can be designed based on known sequences of that class in order to probe genome sequence databases and/or DNA expression libraries derived from samples of DNA purified from extremophile communities. Specific genomes or extremophilic DNA samples can be targeted when searching for a desired characteristic. For example, when looking for enzymes with potential for application in the detergent industry, the alkaline condition employed would necessitate the screening of alkaliphilic organisms as followed by Ito *et al.*, (1998) in the production of detergent enzymes.

To facilitate the genome searches, efficient sequence alignment algorithms must be used. Programs such as BLAST are available (Pearson and Lipman, 1988). However, for detecting more subtle homologues that can be present between members of a protein superfamily, another database such as IDENTIFY (Nevill-Manning *et al.*, 1998) that consists entirely of sequence motifs should be employed to assist in assigning putative protein functions for unknown ORFs. Both approaches will dispense with the need to culture homogeneous samples. However, they are also limited in that they may miss detecting truly unique and unusual proteins by approaching the search based on proteins of known sequences that function under normal environmental conditions. This is clearly highlighted by the archaeal genomes available in which at least 50% of the ORFs cannot be assigned any known function (Bult *et al.*, 1996; Fitzgibbon *et al.*, 1997).

1.4.3.2 Isolating novel enzymes

Novel *in vivo* biotransformations or biodegradation of environmental pollutants have lessened the requirement to isolate specific extremozymes, and there are now several reports on the use of mixed cultures of thermophilic anaerobes that can transform synthetic products through their complete breakdown or via dehalogenation (Maloney *et al.*, 1997a; Maloney *et al.*, 1997b). However, this area has great potential for the establishment of cell-free bioconversions, by isolating and using the purified enzymes responsible for these biodegradations. In addition, it may also be possible to carry out *in vivo* biotransformations to yield a desired end-product. For example, this is shown in the exploitation of the *Pseudomonas* species by the jeans industry for the conversion of indole to indoxyl (a reactive compound that undergoes oxidative dimerization to form indigo) (Ensley *et al.*, 1982; Ensley *et al.*, 1983).

Another potential use of native extremophilic cultures, which does not require a specific extremozyme target is their potential as candidates for novel lead drugs, as they represent vast stores of unique compounds that may inhibit a targeted enzyme or hormone.

1.4.3.3 Enhancing existing enzymes

Another approach to obtaining novel enzymes is to use the current knowledge on the structural basis of stability and activity collated by studying proteins derived from extremophiles. This knowledge may then provide a rational basis for engineering stability into mesophilic enzymes by site-directed mutagenesis. However, the challenge involved here is to produce mutants that can retain full catalytic activity at the enhanced stabilities, which are invariably found to compromise each other (Stoichet *et al.*, 1995). However, a successful example is shown with the engineering of the moderately thermophilic thermolysin-like protease (tlp) from *Bacillus stearothermopholis.* In depth knowledge of the structural features and unfolding processes of the tlp protein family were used to produce a hyperthermostable tlp protein that was able to function at 100°C, in the presence in denaturing agents without forfeiture of enzymatic performance (Vandenburg *et al.*, 1998).

Yet, as shown here, genetic engineering is somewhat limited by the requirement of information on the amino acid sequences and 3D structures of related proteins across wide-, extreme-ranging conditions. Hence, the increasing interest in the approaches that require only the gene of interest and a suitable functional assay for the desired protein. These include:

'forced' evolution, directed evolution, DNA shuffling (also known as molecular breeding) and phage display.

'Forced' evolution has been successfully carried out by Akanuma *et al.* (1998) to improve the thermostability of a mesophilic 3-isopropylmalate dehydrogenase. This involved a step-wise process of random *in vitro* mutagenesis followed by *in vivo* selection in a thermophilic host. Despite its success, it is limited by the requirement for an extremophilic expression host, that is also a minus strain for the gene of interest.

Directed evolution is a method that involves sequential stages of random mutagenesis, cloning and expression in a convenient host, followed by screening for mutants with desired characteristics. A recent review covers many applications of this approach, which can produce changes in the specificity, stability, activity and expression of a host of enzymes (Arnold and Volkov, 1999).

The DNA shuffling first established by Stemmer (1994a, 1994b) involves random recombination of fragments from closely-related gene sequences to create novel genes. This technique was first applied to mutant genes created by error-prone PCR to yield enzymes with enhanced activity towards non-natural substrates (Moore *et al.*, 1997). It has more recently been applied to "family shuffling" an example being the homologous genes coding for cephalosporinases, which were shuffled to create a library of chimeric proteins, thus providing functional diversity (Crameri *et al.*, 1998).

Phage display technology was originally developed to identify and isolate protein domains that could bind tightly to other molecules but has recently been suggested to be a useful new tool for facilitating the discovery of new enzymatic activities from expression libraries (Perdersen *et al.*, 1998). Its sucess is due to the ease of selection of a phage particle, which is dependent on the "display" of target proteins or protein domains on the surface of the phage. The ability of the displayed protein to bind sufficiently tightly to an immobilised substance means that it can be subjected to further rounds of mutagenesis to enhance further the desired characteristic.

1.4.4 SUCCESSFUL AREAS OF EXTREMOZYME APPLICATION

1.4.4.1 Companies incorporating extremophile technology

The soaring interest in extremozymes and their potential to replace chemical catalysts in the manufacture of chemicals, pharmaceuticals, textiles, paper, foods and agricultural chemicals, as well as their ability to participate in bioremediation, is reflected in the number of extremophile programs being developed in several major companies (see Table 1.4)

Company	Program
Altus Biologics (Cambridge, MA, USA)	Artificial extremozymes; cross-linked enzyme crystals
Amersham Pharmacia Biotech (Cleveland, OH, USA)	Thermostable alkaline phosphatase
Archaenzyme (Jerusalem, Israel)	Extremozymes from *Archaea*
Diversa (San Diego, CA, USA)	Extremophile expression libraries
DuPont Merck (Nutley, NJ, USA)	Protease stabilisation in oxidising, alkaline, and hot environments
Genencor International (Rochester, NY, USA)	Protease stabilisation in oxidising, alkaline, and hot environments
Hoffman-La Roche (Basel, Switzerland)	Thermostable Taq polymerase for PCR
Maxygen (Santa Clara, CA, USA)	DNA shuffling of extremozyme genes
Novo Nordisk Biotech (Davis, CA, USA)	Extremozyme cloning and expression
Pfizer Central Research (Groton, CT, USA)	Extremophile screening

Table 1.4 Selected companies incorporating extremophile technology. Reproduced from Persidis (1998).

1.4.4.2 Food, textiles and paper industries

The extracellular-polymer-degrading enzymes from thermo- and hyperthermophiles have many applications in starch and food processing, as well as in the manufacture of detergents and paper.

The complex structure of starch can be broken down by a variety of endo-, exo- and debranching enzymes to yield glucose monomers. The last step in this process is the isomerization of glucose by xylose (glucose) isomerase to produce high-fructose corn syrup, a sweetener usually added to soft drinks. Efforts to improve the efficiency of the process required it to reach temperatures of 100°C. Therefore, research into isolating a more thermostable xylose (glucose) isomerase yielded an enzyme from *Thermatoga maritima* that is stable up to 100°C, with a half-life of 10 min at 115°C (Brown *et al.*, 1993). Other examples in this field are

the development of thermostable xylanases that negate the further use of chlorine in the paper bleaching process (Viikari *et al.*, 1994) and cold-temperature lipases and proteases that can modify the flavour and tenderize meat at temperatures that would also preserve its quality.

1.4.4.3 Chemical industry

The applications of thermostable, organic solvent (low-water) tolerant enzymes that can form complex carbohydrates or optically active compounds are numerous and highly sought after (Wang *et al.*, 1995; McCoy, 1999; de Bont, 1998; Sellek and Chaudhuri, 1999). Selected interest has also focused on alcohol dehydrogenases, due to their ability to oxidise secondary alcohols (Sund and Theorell, 1963) and reduce aromatic ketones to their respective chiral alcohols (Rella *et al.*, 1987), esterases, due to their promotion of transesterification in organic solvents (Luthi *et al*, 1990) and aldolases due to their potential in stereospecific carbohydrate synthesis (Wang *et al*, 1995).

1.4.4.4 Pharmaceutical and scientific industries

A major scientific advancement was the replacement in the PCR process of the Klenow fragment of DNA polymerase with the thermostable Taq polymerase from the thermophilic eubacterium *Thermus aquaticus* (Chien *et al.*, 1976). Recent discoveries have seen the production of several DNA polymerases with higher proof-reading capabilities, such as *Pwo* pol (Frey and Suppmann, 1995). In addition, the methodss described in 1.4.3.3 are being employed to screen extremophilic enzymes to produce novel drug compounds that are enantiomerically pure (Potera, 1998).

1.4.4.5 Environmental biotechnology

Extremozymes employed in this field are able to benefit mankind directly through the procurement of fuel sources, as well as having a potentially tremendous effect on the environment through the degradation of toxic waste products.

Oil-field commercial application of high-pressure, high-pH and high-temperature enzymes is being used to enhance the flow of oil or gas in drilling operations, in which they operate by degrading materials, reducing the viscosity and enabling oil to flow through fissures (Trebbau de Acevedo and McInerny, 1996; McCutchen *et al.*, 1995).

The discovery of highly thermostable extracellular-polymer degrading enzymes has also been exploited in an *in vitro* enzymatic pathway established by Woodward *et al.* (1996) to

produce molecular hydrogen, which employed thermostable cellulases as part of the *in vitro* pathway to hydrolyze cellulose to glucose. An alternative cellulosic waste product, maple tree sap, was also converted to hydrogen using an *in vitro* pathway. This included the following thermophilic enzymes, an invertase, glucose isomerase, glucose dehydrogenase and a hydrogenase (Woodward and Orr, 1997). The basis of the *in vitro* method of hydrogen production using these pathways (Woodward *et al.*, 1996) is the employment a thermophilic nicotinamide-dependent glucose dehydrogenase isolated from *T. acidophilum*, which grows optimally at 59°C (Bright *et al.*, 1993). The remaining portion of the pathway involves a hyperthermophilic nicotinamide-dependent hydrogenase isolated from *Pyrococcus furiosus*, which grows optimally at 100°C (Bryant and Adams, 1989). Glucose derived from cellulosic waste is oxidized to glucono-δ-lactone by glucose dehydrogenase, which immediately hydrolyses to gluconic acid (Harrison, 1931). The cofactor, $NADP^+$, is simultaneously reduced. The presence of a hydrogenase in this system which is able to interact with this physiological electron carrier (Egerer *et al.*, 1982; Bryant and Adams, 1989), brings about the regeneration and recycling of $NADP^+$ with the concomitant production of molecular hydrogen (Fig. 1.6). Although this pairing has resulted in the successful production of stoichiometric yields of hydrogen from glucose, the temperature limitation of the GDH enzyme means that this reaction should be carried out around its optimal temperature in order to prevent thermal inactivation, thus causing the hydrogenase to operate below its temperature optimum. Additional studies have shown that *T. acidophilum* GDH is inhibited by levels of glucose exceeding 50 mM (Robinson, 1995). These factors are believed to affect the efficiency of turn over of hydrogen. Therefore it was considered necessary to search for another thermophilic pyridine-dependent GDH with a higher operating temperature than the *T. acidophilum* GDH and, which would also not undergo substrate inhibition, in order to partner the hyperthermophilic hydrogenase in this system. This pathway is considered to be very advantageous as it produces molecular hydrogen as the end product without the generation any of pollutants. (Participation in this pathway of the enzyme produced in this work is described in detail in chapter 7).

The ability of the bacterium *Deinococcus radiodurans* to survive acute exposures to ionising radiation without lethality (Daly *et al.*, 1994) was exploited by engineering it to express foreign genes involved in the transformation of organopollutants (Lange *et al.*, 1998).

Finally, in reponse to the public concern about the harmful effects of petrochemical-derived plastic materials in the environment, manufacturers are developing biodegradable plastic materials. These plastics are produced by microbes as an energy reserve material. They are called polyhydroxyalkonaoates (PHAs) and are polyesters of various hydroxyalkonoates (Lee, 1996). One of the PHAs currently being marketed by the company Monsanto is 'Biopol plastic', which combines polyhydroxybutyrate (PHB) (a brittle homopolymer) with polyhydroxvalerate (PHV) (a nonbrittle copolymer) to create the end-product, poly(3-hydroxybutyrate-co-3-hydroxyvalerate) (PHBV) (Kim *et al.*, 1998). However, the ability of several halophilic archaea to produce these copolymers could also be exploited, as suggested by Cowan (1992). This would have the advantage of lower production and recovery costs, as well as generating different PHAs in response to altered organic acid substrate supplies.

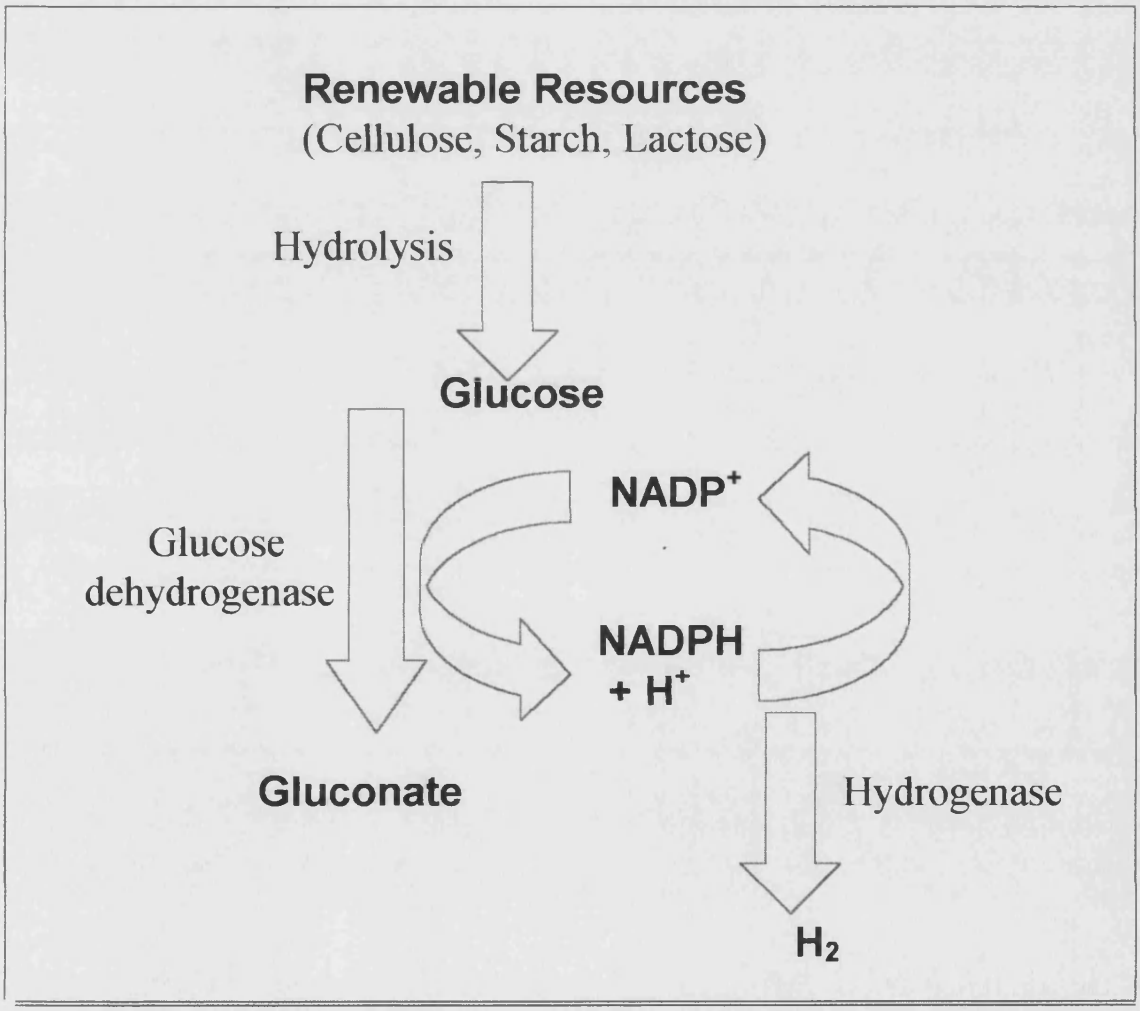


Fig. 1.6 **Diagram depicting the enzymatic pathway for the conversion of renewable resources to hydrogen.**

1.5 DEHYDROGENASES AND GLUCOSE DEHYDROGENASE

1.5.1 GLUCOSE DEHYDROGENASES

Two types of glucose dehydrogenases (GDH) have been described: pyrroloquinone-quinone (PQQ)-dependent (EC 1.1.99.17) and pyridine-nucleotide-dependent GDHs (EC 1.1.1.47). PQQ-dependent enzymes seem to be restricted to some Gram-negative bacteria, such as several species of *Gluconobacter*, *Acinetobacter*, *Klebsiella*, *Pseudomonas* and *Escherichia* (Cleton-Jansen *et al.*, 1988; Cleton-Jansen *et al.*, 1989 and literature cited therein). Also according to primary structure analysis, both GDH-types are not related (Cleton-Jansen *et al.*, 1988). Pyridine-nucleotide-dependent GDHs have been isolated from all three domains of life: Bacteria, Archaea and Eukarya. These enzymes are found to be typically homomeric tetramers that exhibit dual cosubstrate specificity for NAD^+ and $NADP^+$ (Campbell *et al.*, 1982; Jany *et al.*, 1984; Giardina *et al.*, 1986; Lampel *et al.*, 1986; Smith *et al.*, 1989). Exceptions are the tetrameric enzymes of *Schizosaccharomyces pombe* (Tsai *et al.*, 1995) and *Halobacterium halobium* (Sonawat *et al.*, 1990), for which a high $NADP^+$ specificity has been described, the dimeric enzymes from *Haloferax mediterranei* (Bonete *et al.*, 1996) and *Thermoproteus tenax* (Siebers *et al.*, 1997) and the monomeric enzyme of *Corynebacterium* species (Giardina *et al.*, 1986).

1.5.2 MEDIUM-CHAIN ALCOHOL/POLYOL DEHYDROGENASE/REDUCTASES (MDR)

As outlined by Jörnvall *et al.* (1984; 1987) and Persson *et al.* (1991), the enzymes characterised as medium-chain (former long-chain, 350 - 375 residues) dehydrogenase/reductases consist of the following: alcohol dehydogenases (and glutathione-dependent formaldehyde dehydrogenases), polyol dehydrogenases, threonine dehydrogenases, archaeal glucose dehydrogenases and eye lens reductase-active ζ-crystallins, and also include *E. coli* quinone oxidoreductase, Torpedo VAT-1 protein, and enoyl reductases of mammalian fatty acid and yeast erythronolide synthases. This protein family forms one of the three deeply branching evolutionary lines within the superfamily of pyridine-nucleotide-dependent alcohol/polyol/sugar dehydrogenases. They are opposed to the branch of the short-chain (about 250 residues), non-zinc-containing dehydrogenases (to which the GDHs of *Bacilli*, the zinc-free

ribitol dehydrogenases and the *Drosophila* alcohol dehydrogenase belong) and to the branch of iron-activated alcohol dehydrogenases. Two representatives of the MDR family that have known structures (Eklund and Bränden, 1976; John *et al.*, 1994) are shown in Figs 1.6a and b. They both exhibit two domains, the central nucleotide-binding domain (≈180 - 300 residues) flanked by catalytic domains. Both of these members also contain a catalytic and structural zinc atom. Both domains are separated by a deep active site cleft that is freely accessible to the solvent (Bränden, 1975; John *et al.*, 1994). The catalytic zinc is at the bottom of the cleft and the lobe containing the structural zinc is at the mouth of the cleft.

(a)

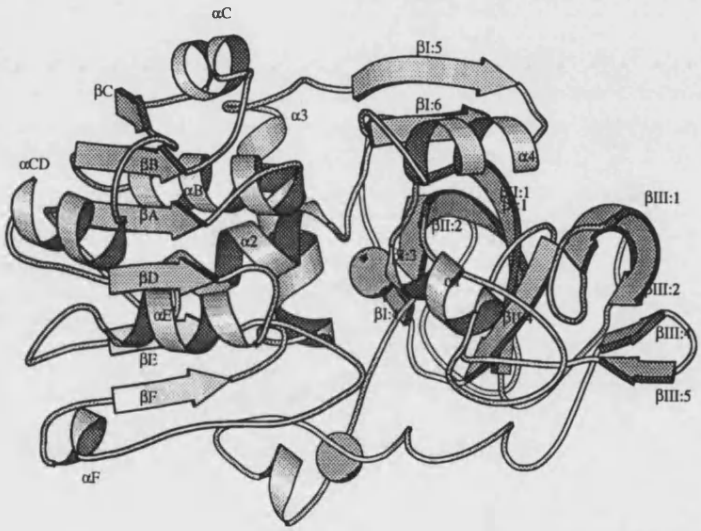
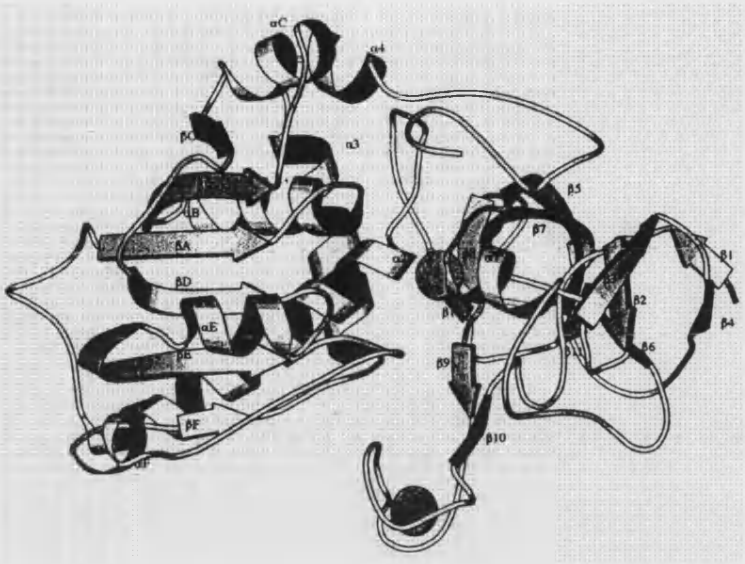


Fig. 1.7 a and b Schematic representations of (a) the LADH (liver alcohol dehydrogenase) monomer and (b) the GDH monomer from the Archaeon *Thermoplasma acidophilum*. Both were produced using MOLSCRIPT (Kraulis, 1991) and were reproduced from Rossjohn (1994) and John *et al.* (1994) respectively.

(a) The secondary structural elements defined by Eklund *et al.* (1976) are: βIII:1, 9-14; βIII:2, 22-29; βII:1, 33-40; βI:1, 41-45; α1, 46-53; βIII:3, 62-65; βI:2, 68-71; βII:2, 72-78; βII:3, 86-87; βI:3, 88-92, βIII:4, 129-132; βIII:5, 135-138; βIII:6, 145-146; βII:4, 148-152; βI:4, 156-160; α2, 168-179; αA, 180-188; βA, 193-199; αB, 201-215; βB, 218-223; αC, 229-236; βC, 238-243; αCD, 250-258; βD, 263-269; αE, 271-280; βE, 287-293; βS, 299-303; 3^{10}S, 304-311; βF, 312-318; α3, 324-388; βI:5, 347-352; α4, 355-366; βI:6, 369-374. Zinc ions are shown as spheres.

(b)



(b) Amino-terminal catalytic domain: β1, residues 2-4; β, 6-9; β3, 13-17; β4 20-22; β5, 30-37; β6, 62-64; β7, 69-76; β8, 87-90; β9, 92-94; β10, 116-118; β11, 132-134; β12, 139-141; α1, 41-47; α2, 154-174. Nucleotide-binding domain: βA, residues 186-191; βB, 209-215; βC, 233-236; βD, 245-250; βE, 265-273; βF, 298-300; αB, 195-205; αC, 221-228; αE, 225-262; αF, 290-294. Carboxy-terminal catalytic domain: α3, residues 308-321. Zinc ions are shown as spheres.

1.5.3 THE NUCLEOTIDE-BINDING FOLD AND COFACTOR BINDING

The nucleotide-binding fold and mode of cofactor binding appear to be structurally conserved throughout the dehydrogenase family (Eklund *et al.*, 1984; Ohlsson *et al.*, 1974). The fold consists of a characteristic αβαβαβ motif known as the 'Rossmann fold' (Eklund *et al.*, 1976; Rossmann *et al.*, 1974; Rao and Rossmann, 1973). However, the additional helix commonly found in members of the MDR family (Persson *et al.*, 1994) is not part of the classical Rossmann fold.

Several structurally-based studies on sequence patterns found the finger-print motif of the nucleotide-binding domain and a residue at the end of strand βB to be involved in cofactor specificity (Wierenga *et al.*, 1985; Wierenga *et al.*, 1986; Hanukoglu and Gutfinger, 1989; Scrutton *et al.*, 1990; Baker *et al.*, 1992a). NAD^+ binding is proposed to require a GxGxxG

sequence motif plus a negatively-charged residue at the end of a βB strand, whereas specificity for $NADP^+$ is achieved by a GxGxxA sequence motif and a positively-charged residue near the end of strand βB. However, observations by Baker *et al.* (1992b) and Edwards *et al.* (1996) have shown that the alanine residue present in the finger motif is not a prerequisite for $NADP^+$ binding and that the Gly/Ala residue motif actually determines the hydrogen-bonding system between adenine ribose and the GxGxxG/A loop (Baker *et al.*, 1992b; Rossjohn, 1994). Therefore, further considerations of sufficient space and the location of a positively-charged residue for the additional adenine ribose phosphate group of the NADPH cofactor are required (Edwards *et al.*, 1996).

The NAD(P) binds in an extended conformation to the C-terminal ends of the β-strands of the Rossmann fold. The adenosine moiety sits in a hydrophobic pocket provided by the βαβ fold. The positioning of the adenosine moiety is structurally conserved (Wierenga *et al.*, 1985), whereas there is more structural deviation in the site of the nicotinamide ring. Dehydrogenases accept hydride ions directly from the substrate in a stereospecific manner. Based on this stereospecificity, dehydrogenases have been divided into two classes, A and B. Members of the MDR family are all found to share A specificity (Schneider-Berlöhr *et al.*, 1986). This is where the pro-R hydrogen of the alcohol is transferred to the A-side of the nicotinamide ring. This is opposed to the GDHs from *Bacillus*, which belong to the short-chain alcohol-polyol dehydrogenase family, which are all B-specific (Jörnvall *et al.*, 1987; Schneider-Bernlöhr *et al.*, 1986).

1.5.4 THE CATALYTIC DOMAIN

Side chain modification studies carried out on horse liver alcohol dehydrogenase – also an MDR enzyme that is thought to share very high structural homology with the GDHs of *T. acidophilum* and therefore possibly with *S. solfataricus* (John *et al.*, 1994; Edwards *et al.*, 1996) – revealed that the presence of pyridine-containing compounds protected a lysine and two arginine residues from modification and were all therefore thought to participate in coenzyme binding (Bränden *et al.*, 1975). However, no decrease in activity was found to occur on modification of histidine and tyrosine residues. In addition, further analysis via a structural comparison of four medium-chain dehydrogenases, quinone oxidoreductase (QOR) from *E. coli*, mammalian reductase-active ζ-Crystallin, alcohol dehydrogenase from horse liver (LADH),

and GDH from *T. acidophilum* (Edwards *et al.*, 1996), revealed a high structural conservation in the region of the active site (helices α1 and α2 and the strand β1). However, the low sequence identity shown by these members also caused significant differences, such as the possession of co-ordinated zinc atoms (not present in the reductases) and the charge-nature of the binding site. A comparison of the residues that line the substrate-binding sites in LADH and GDH and their equivalents in QOR and ζ-crystallin also showed marked differences (see Table 1.5). However, despite the high degree of amino acid sequence divergence, a stable enzyme fold has been retained within this family, thus demonstrating their evolution from a common ancestor. Also the differences observed in the catalytic domain allowed the evolution of new enzyme activities.

QOR	ζ-Crystallin	LADH	GDH
-	-	Zn	Zn
Asn 41	Asn 48	Cys 46[b]	Cys 40[b]
Thr 63	Thr 71	His 67[b]	His 67[b]
Leu 123	Ile 131	Cys 174[b]	Glu 155[b]
-	-	Leu 116[c]	-
-	-	Phe 93	Val 92
-	-	Leu 141[c]	-
-	-	Phe 140[c]	-
Ile 43	Val 50	Ser 48[d]	Thr 42[d]
Asn 240	Cys 248	Val 294	Thr 276
Tyr 52	-	Leu 57	-
Ser 265	Ser 270	Ile 318	Ser 303
Leu 266	Leu 271	Phe 319[e]	Val 304[e]
Arg 263	Gly 268	Gly 316[e]	Ala 301[e]

Table 1.5 Residues lining the substrate-binding sites for the MDR enzymes: QOR[a], (quinone oxidoreductase from *E. coli*), mammalian reductase-active ζ-Crystallin ,LADH (alcohol dehydrogenase from horse liver), and GDH from *T. acidophilum*. Reproduced from Edwards *et al.*, 1996.

Note. A dash denotes no equivalent residue.

[a] The potential substrate-binding site for QOR has been identified based on the X-ray structure (Thorn *et al.*, 1995), analogy to LADH, and biochemical considerations.

[b] Residues that bind the catalytic zinc in LADH and GDH.

[c] Residues that are part of the structural zinc-binding loop in LADH.

[d] Catalytic residues in LADH and GDH but not part of substrate-binding site.

1.5.5 PUTATIVE MECHANISM OF LIVER ALCOHOL DEHYDROGENASE

Firstly, a conformational change is seen in *T. acidophilum* GDH on binding cofactor (Rossjohn, 1994), as in many other dehydrogenases. This was determined using fluorescence and inhibition studies, which show that *T. acidophilum* GDH follows a compulsory ordered mechanism, in which cofactor binds first and then undergoes transition to the closed state. However, further analysis by Cedergren-Zeppenzauer (1983), has opposed this hypothesis and has showed that the ternary complex of LADH with NADH and imidazole still retained an open conformation. So consequently, not only cofactor binding, but also the transition to the closed state is needed for the creation of a steric strain in this region of the polypeptide chain.

The actual enzyme mechanism relies on the presence of a Zn^{2+} ion, tetrahedrally ligated by two cysteine residues, a histidine and a water molecule (Eklund *et al.*, 1976) or a hydroxyl ion, depending on the pH (Bränden *et al.*, 1975). Recent work to deduce the mechanism of the biological oxidation of alcohols to aldehydes and ketones by LADH (Kimblin *et al.*, 1997; Bergquist and Parkin, 1999) has proposed the involvement of a four co-ordinate zinc hydroxide complex. This complex is capable of generating a four co-ordinate terminal alkoxide derivative from an un-activated alcohol such as ethanol (Fig. 1.8). Subjection of the zinc co-ordinated complex to alkaline conditions promotes the ionisation of the zinc co-ordinated water and consequently makes it more reactive. This theory has been augmented by the work of Groves and Olsen (1985), that has shown that zinc co-ordinated water can have a pK_a as low as 7 (i.e. be hydroxide-like or nucleophilic at optimum pH and therefore become a highly potent nucleophile at alkaline pH values). Berquist and Parkin (1999) also showed that these alkoxide intermediates are sufficiently activated to be capable of transferring "a hydride" to a substrate. However, Bränden (1977) proposed a mechanism in LADH that also required the involvement of the hydrogen bond system of Ser 48 and His 51 in hydride transfer. Yet more recent work on enzyme catalysis using a thermophilic alcohol dehydrogenase (Kohen *et al.*, 1999) has also proposed the involvement of enzyme dynamics and hydrogen tunnelling in enzyme-catalysed hydrogen transfer reactions, where a greater role for the mobility of the enzyme in C-H bond cleavage at higher temperatures is suggested.

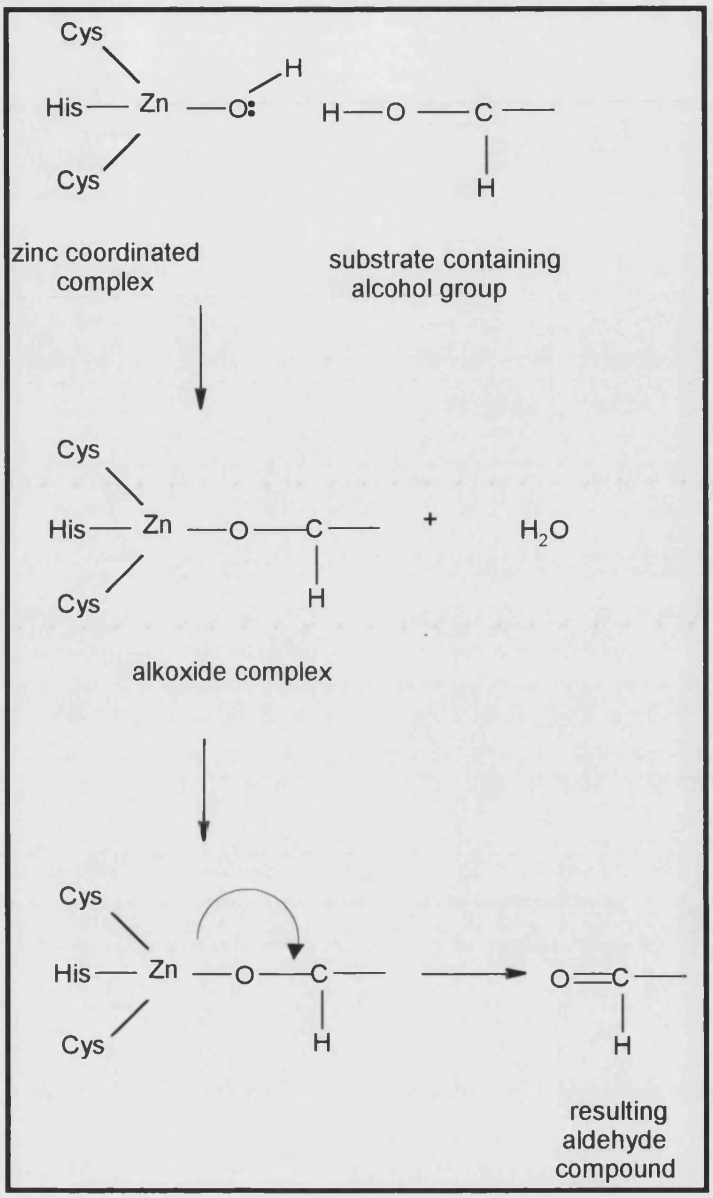


Fig. 1.8 Putative mechanism of catalysis of alcohol oxidation via a four co-ordinate zinc hydroxide complex to form a four co-ordinate alkoxide derivative [adapted from Bergquist and Parkin, (1999)].

The proposed reaction mechanism of glucose oxidation by glucose dehydrogenase is similar to that proposed for the liver alcohol dehydrogenase, for reasons of strong homology displayed by the possession of two co-ordinated zinc atoms and membership of the MDR family and also due to the similarity of their reactions which essentially catalyse the conversion of an alcohol group to a carbonyl. The formation of D-Glucono-1,5-lactone (Fig. 1.9) occurs via

the release of two hydrogen atoms from C1 of D-Glucopyranose and the concomitant reduction of $NAD(P)^+$. The oxidation of D-Glucopyranose causes the production of another pyranose compound with a carbonyl group at C1 (D-Glucono-, 5-lactone). However, this compound is short-lived in solution and is immediately hydrolysed by water to form D-Gluconic acid, which possesses a linear structure. The spontaneous hydrolysis of gluconolactone, which is facilitated by the alkaline conditions, also displaces the direction of equilibrium even further to the right.

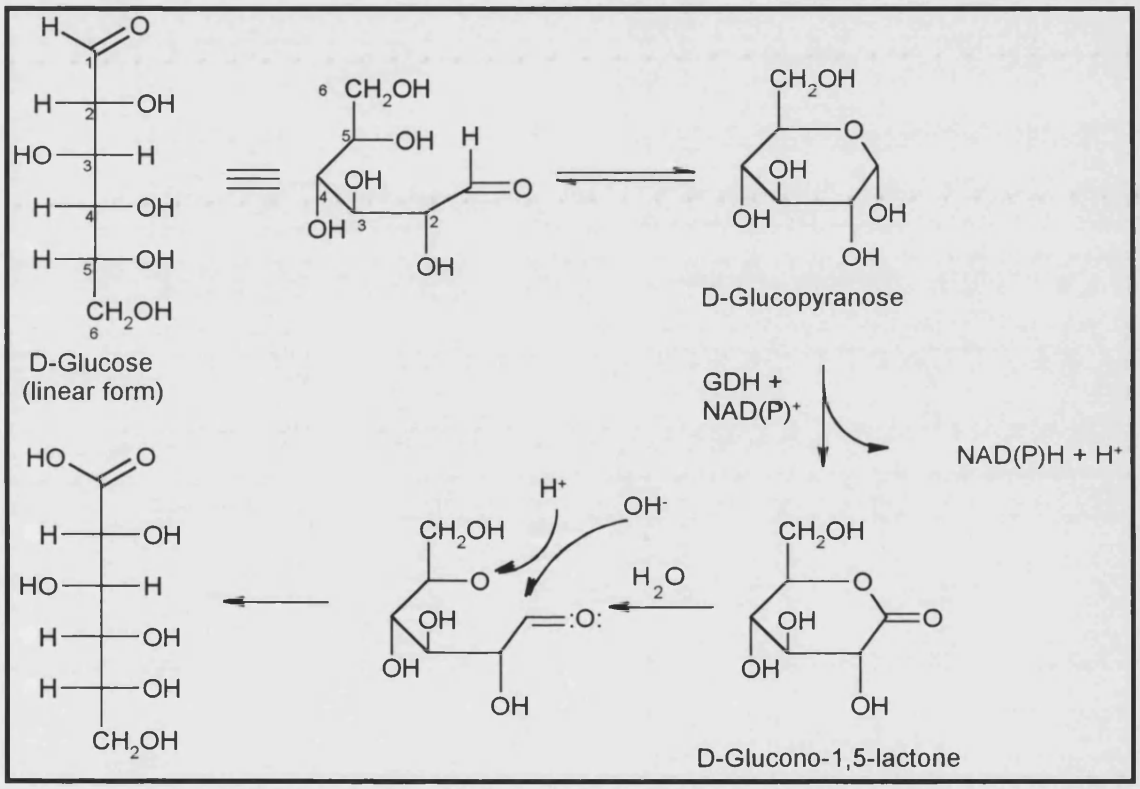


Fig. 1.9 **Reaction mechanism depicting the conversion of D-Glucose to D-Gluconic acid.**

1.6 AIMS

The aims of the project were to isolate, clone and sequence the glucose dehydrogenase gene from *Sulfolobus solfataricus*. The gene was then to be expressed in the mesophilic host, *Escherichia coli*, and the recombinant GDH to be purified to homogeneity. The aim was then to determine the characteristic properties of the GDH protein to possibly gain

some insight into how it has adapted to function under such extreme conditions. Comparison of these properties with enzymes of the MDR family, in particular the GDH from the Archaeon *Thermoplasma acidophilum*, would confirm its evolutionary grouping and highlight some common structural features. Although there might not be sufficient time in which to obtain a three-dimensional crystal structure, it was intended to set up preliminary crystallisation trials and in addition, construct a 3D homology model of the *S. solfataricus* enzyme using the crystallographically-determined atomic structure of the *T. acidophilum* enzyme as a base. However, the main aim of this work was to employ the *S. solfataricus* enzyme in the *in vitro* enzymatic hydrogen production system and compare the yields and rates of hydrogen production with the system employing the *T. acidophilum* enzyme.

2 Chapter Two

General Materials & Methods

2.1 MATERIALS

All reagents were of analytical grade or of the finest grade commercially available.

2.1.1 CELLCULTURE AND PROTEIN PREPARATION

Yeast extract and bacto-agar were supplied by Fischer Scientific, Loughbrough, Leicestershire, U.K. Tryptone was purchased from Amersham, Bury, England. Ampicillin (sodium salt) and Carbenicillin were obtained from Sigma, UK. Sonication was carried out using a MSE 150 watt Ultrasonic Disintegrator Mk2 from, MSE Scientific instruments, Crawley, Sussex, U.K.

2.1.2 MOLECULAR BIOLOGY

All details of molecular biological materials and techniques are supplied in chapter 4.

2.1.3 PLASMIDS AND ORGANISMS

E.coli strain JM109 was supplied by Promega, Madison, WI, U.S.A. *E.coli* strain BL21 (DE3) was obtained from Novagen, Abingdon, U.K.

2.1.4 CHROMATOGRAPHY

The HiLoad 16/60 Superdex 200 prep grade 30pg gelfiltration column and the Fast Protein Liquid Chromatography (FPLC) equipment were from Pharmacia Biotech, Uppsala, Sweden. Matrex gel Red A was purchased from Millipore Ltd., Watford, Hertfordshire, U.K. The Affinity Cibacron blue F3GA dye 5ml column and the Poly-Prep columns which were used to make 2ml Red gel A columns came from Bio-Rad Laboratories, Hercules, California, U.S.A.

2.1.5 ENZYME ASSAYS

NAD^+ , $NADP^+$, $MgCl_2$ and D-Xylose were obtained from Sigma-Aldrich Ltd., Poole, Dorset, U.K. Glucose was from Fischer Scientific, Loughbrough, Leicestershire, U.K. The Spectrophotomters used were :- Lamda 3B UV/Vis and Lamda Bio 230V UV/Vis/NIR from Perkin-Elmer Corporation, Norwalk, Connecticut U.S.A

2.1.6 ELECTROPHORESIS

Protogel™ was purchased from National Diagnostics (U.S.A). TEMED and ethidium

bromide solution were obtained from Sigma. SeaKem LE agarose was obtained from FMC BioProducts, Rockland, Maine, U.S.A. and Low Melting Point Agarose was from Schwarz/Mann Biotech, Cleveland, Ohio, U.S.A. Protein low and broad range molecular weight standards were obtained from Bio-Rad Laboratories Ltd. 1kB Ladder was from GIBCO BRL Life Technologies, New York, U.S.A. Protein Electrophoresis equipment was from Atto-Corp., Mini-Atto System.

2.1.7 OTHER GENERAL REAGENTS

Tris was from Boehringer, Mannheim, Germany. MES was from Sigma-Aldrich Company Ltd. Poole, Dorset, U.K.

2.2 BIOCHEMICAL TECHNIQUES

2.2.1 Glucose Dehydrogenase Assay

Glucose dehydrogenase activity was determined spectrophotometrically by following the change in absorbance at 340 nm, corresponding to the reduction of $NADP^+$. The standard 1ml assay mixture contained 0.5 mM $NADP^+$, 5 mM D-glucose and in 50 mM Tris/HCl pH 9, 20 mM $MgCl_2$ at 70°C.

2.2.2 Protein Estimation

Protein concentration was determined spectrophotometrically using the method of Bradford (1976), by comparison with a standard curve of absorbance at 595 nm vs. protein concentration over a protein concentration range of 0-20 μg bovine serum albumin/ml.

2.2.3 Preparation of Crude Cell Extract

8g of cell paste was re-suspended in 10 ml 50 mM Tris/HCl at various pH values or 50 mM Mes at pH 5.5, with 20 mM $MgCl_2$ and 1mM PMSF. Cells were broken by five 30s bursts of sonication at an amplitude of 9 micrometers with 30s rest periods on ice.

The cell debris was then pelleted out by centrifugation at 18,000 x g for 1h at 4°C in a Sorvall SS-34 rotor.

2.2.4 FPLC Superdex 200 Gel Filtration

2 ml of protein sample was loaded in 50 mM Tris/HCl, 20 mM $MgCl_2$ of pH values 7-9 or 50 mM Mes, 20 mM $MgCl_2$ pH 5.5 buffer onto the FPLC HiLoad 16/60 Superdex 200 prep grade gel filtration column at a flow rate of 1 ml/min; this flow rate was used through out 1 ml fractions were collected after the 20 ml void volume.

2.2.5 FPLC HiTrapQ anion exchange

Samples containing up to 30 mg of total protein in 50 mM Tris/HCl, 20 mM $MgCl_2$ pH 8.5 buffer were loaded on to the anion exchange column at a flow rate of 1 ml/min; this flow rate was used through out (as will also apply to the following FPLC columns used).The loaded column was

washed with 15 ml of 50 mM Tris/HCl, 20 mM $MgCl_2$, pH 8.5 buffer and the bound protein was then eluted using a 25 ml 0-1 M sodium chloride gradient in the same buffer.

2.3.6 FPLC HiTrapSP cation exchange

Samples containing up to 30 mg of total protein in 50 mM Mes, 20 mM $MgCl_2$ pH 5.5 buffer were loaded on to the cation exchange column at a flow rate of 1 ml/min. The washing and elution steps were carried out as in 2.3.5.

2.3.7 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Discontinuous SDS-PAGE was performed according to the method of Laemmli (1970). The gels were fixed and stained for 1 - 2h in Fix solution (aqueous 45% (v/v) methanol / 10% (v/v) acetic acid), containing 0.25% (w/v) Coomassie Blue, and destained in Fix solution (5% (v/v) methanol and 7.5% (v/v) acetic acid).

3 Chapter Three

Purification of native Glucose dehydrogenase

3.1 INTRODUCTION

Nicotinamide-nucleotide-dependent glucose dehydrogenase (EC 1.1.1.47) has been isolated from a variety of different sources, such as mammalian liver (Campbell *et al.*, 1982), and a variety of eubacteria including *Bacillus megaterium* (Pauly and Pfleiderer, 1975) and *Bacillus subtilis* (Lampel *et al.*, 1986). More notable are the archaeal sources of GDH, namely *Haloferax mediterranei* (Bonete *et al.*, 1996), *Thermoplasma acidophilum* (Smith *et al.*, 1989), *Thermoproteus tenax (*Siebers *et al.*, 1997) and, most salient to the subject of the current thesis, is the previously-published work on the purification and characterisation of GDH from *Sulfolobus solfataricus* (Giardina *et al.*, 1986). *

The various purification strategies employed to isolate the GDH enzymes from their respective hosts showed agreement on the successful use of dye affinity chromatography (Table 3.1) (Giardina *et al.* 1986; Smith *et al.*, 1989; Bonete *et al.*, 1996). This method relies on the ability of the covalently-bound dye ligands to mimic nucleotide containing compounds, for example ATP and $NAD(P)^+$ (Garg *et al.* 1996). Utilising this method, a high level of purification was attained for the *T. acidophilum* GDH, which is thought to share high sequence homology with the *S. solfataricus* GDH. This is based on evidence of the high sequence identity (58%) of another metabolic enzyme, citrate synthase, isolated from these two species (Connaris *et al.*, 1998) and also their identical methods of hexose metabolism.

This chapter describes various purification strategies employed to purify the native GDH enzyme and obtain its N-terminal sequence. Procurement of the N-terminal sequence would allow the design of a probe for hybridisation experiments and hence lead to the acquisition of the GDH gene.

* It must be noted that this publication employed an alternative strain of *S. solfataricus* (MT-4) as opposed to the strain (P1) used in this work.

Glucose dehydrogenase from *Sulfolobus solfataricus*, (Giardina *et al*, 1986)

Step	Total activity (U)	Protein (mg)	Specific activity (U/mg)	Yield %	Fold Purification
Extraction and $(NH_4)SO_4$ precipitation	2540	3620	0.7	100	1.0
SP-Sephadex	1752	325	5.4	70	7.8
Phenyl-Sepharose	1480	81	18.2	59	26.4
Affi-Gel Blue	699	1.6	437	28	623.0

Glucose dehydrogenase from *Thermoplasma acidophilium*, (Smith *et al*, 1989)

Step	Total activity (U)	Total Protein (mg)	Specific activity (U/mg)	Yield %	Fold purification
Cell extract	160	2080	0.08	100	
Methanol extraction	130	800	0.16	81	2
Mono Q chromato-focusing	99	20	5.0	62	63
Red Gel A	80	2.8	28.6	50	358
Mono P chromato-focusing	54	-	-	34	-
Superose 12 Gel filtration	32	0.1	320	20	4000

Glucose dehydrogenase from *Haloferax mediterranei*, Bonete *et al.*, 1996

Step	Total protein (mg)	Specific activity (U/mg)	Yield %	Purification Fold
Cell extract	790	0.7	100	-
Sepharose-4B	25	17	72	23
DEAE-cellulose	23	17	66	23
Blue-Sepharose	1	190	33	271
Red-Sepharose	0.2	550	19	786

Table 3.1 Previous purification methods employed to isolate archaeal glucose dehydrogenases.

3.2 MATERIALS AND METHODS

3.2.1 MATERIALS

All reagents were of analytical grade or of the finest grade commercially available.

3.2.1.1 Organisms

Sulfolobus solfataricus strain P1 was supplied as cell paste from Prof. R. Sharp, CAMR (Centre for Applied Microbiological Research), Porton Down, Wilts, U.K.

3.2.1.2 Reagents

MES, xylose, nitroblue tetrazolium and phenazine methosulfate were from Sigma-Aldrich Company Ltd. Poole, Dorset, U.K

3.2.1.3 Equipment

The Semi-dry transfer Blotter was from Pharmacia Biotech, Uppsala, Sweden and Immobilon™-P Transfer membrane was from Millipore Corp., Bedford, UK.Matrex Gel Red A was purchased from Millipore Ltd., Watford, Hertfordshire, U.K. For all other equipment cited, refer to Chapter 2.

3.2.2 METHODS

Methods outlining the following procedures are described in detail in Chapter 2, section 2.3: Preparation of cell extract, FPLC HiLoad 16/60 Superdex 200 prep grade gel filtration, FPLC HiTrapQ anion exchange, FPLC HiTrapSP cation exchange.

3.2.2.1 Matrex Gel Red A Chromatography

Preliminary studies were carried out with a 2 ml column prepared by pouring Matrex Gel Red A into a Poly-prep column. The column was equilibrated with 10 ml of 50 mM MES pH 5.5, 20 mM $MgCl_2$, *S. solfataricus* cell extract samples in the same buffer were applied and the column was also washed in 10 - 15 ml of the same buffer. Elution of GDH was achieved using 5 ml of 50 mM xylose or glucose plus 0.5 mM $NADP^+$ in the same buffer or alternatively with 1 M sodium chloride in the same buffer. A third elution approach employed 50 mM Tris/HCl pH 8.5, 20 mM $MgCl_2$.

An attempt was made to scale up this purification step in which a 30 ml Matrex Gel Red A column used in conjunction with a Gradifrac fraction collector, was loaded with protein in

50 mM MES pH 5.5, 20 mM $MgCl_2$, at a flow rate of 1 ml/min. The loaded column was washed with 90 ml of the same buffer and the bound protein was eluted using 0 - 1.5 M sodium chloride gradient in the same buffer.

3.2.2.2 F3GA Cibacron Blue Chromatography

Samples containing up to 50 mg of total protein in 50 mM Tris/HCl pH 7.0, 20 mM $MgCl_2$, were loaded on to the affinity column at a flow rate of 1 ml/min. The loaded column was washed in 15 ml of the same buffer and the bound protein was eluted using 10 ml of 50 mM xylose and 0.5 mM $NADP^+$ in the same buffer.

3.2.2.3 Native Polyacrylamide Gel Electrophoresis

Native PAGE was performed as SDS-PAGE (2.3.7) but omitting SDS from loading the buffer and gel buffer concentrates.

3.2.2.4 Western blotting

The SDS gel, PVDF membrane and filter paper were pre-soaked in Western transfer buffer (25 mM Tris/HCl, 192 mM glycine, pH 8.3, 20% (v/v) methanol) for 10 min. The gel and the membrane were sandwiched between the filter paper, and the protein was transferred from the gel with a current of 250 mA for approximately 30 min. The protein was fixed and stained in Stain solution (0.1% (w/v) Coomassie, 50% (v/v) methanol), and 50 - 100% (v/v) methanol was used to destain the blot.

3.2.2.5 Enzyme activity stain

Following electrophoretic separation in a native discontinous gel system, the gel was covered in 50 ml of freshly prepared staining solution containing 0.65 mM $NADP^+$, 0.4 mM nitro blue tetrazolium, 0.07 mM phenazine methosulphate, 100 mM glucose and 40 mM Tris/HCl, pH 8.0. The gel was incubated in the dark at 37°C (and also at 70°C) until a blue-brown band appeared. The stained bands were cut out and run on SDS PAGE.

3.2.2.6 Washing procedure of SDS-PAGE

Following SDS electrophoresis as described in 2.3.7, but with the omission of the Coomassie Brilliant Blue staining, the gel was washed for 2 h in 10 mM sodium phosphate buffer, pH 7.1, to remove the SDS. The buffer was changed every half hour. The activity stain (described in 3.2.2.5) was then performed.

3.2.2.7 N-terminal micro-sequencing

N-terminal micro-sequencing was performed by Mrs Janice Young at Zeneca Pharmaceuticals using an Applied Biosystems 477 Sequencer (modified to run blot cycles) with a gas phase TFA delivery.

3.3 RESULTS

3.3.1 PURIFICATION OF *SULFOLOBUS SOLFATARICUS* GDH

Five different purification strategies were attempted and the data are summarised in Table 3.2 . The samples purified using these approaches were viewed on SDS-PAGE as shown in Figs 3.1 - 3.5.

APPROACH	STEP	BUFFER pH	VOLUME (ml)	TOTAL ACTIVITY (µmol/min/ml)	TOTAL PROTEIN (mg)	SPECIFIC ACTIVITY (µmol/ml/min/mg)	FOLD PURIFICATION	YIELD (%)
	CE	8.0	2.0	11.0	35.2	0.313	-	100
1	RG	8.0	3.0	6.5	1.2	5.42	17.7	58.9
	GF	8.0	1.0	0.15	0.013	11.45	37	1.37
	CE	7.0	2.0	9.4	38.8	0.242	-	100
2	GF	7.0	6.0	10.2	8.4	1.2	4.96	109
	RG	7.0	0.5	0.214	0.025	8.56	35.4	2.3
	CE	5.5	2.0	28.2	85.0	0.33	-	100
3	GF	5.5	7.0	29.2	31.5	0.93	2.8	104
	H-SP	5.5	2.5	6.04	0.81	7.45	22.6	69
	RG	5.5	1.0	5.82	0.142	41.0	124	20.6
	CE	5.5	4.0	35.0	107.0	0.33	-	100
4	GF	5.5	20	43.7	45.5	0.96	2.9	125
	RG	5.5	3.5	3.63	0.1	36.3	110.0	10.4
	CE	7.0	2.0	19.4	96.8	0.2	-	100
5	GF	7.0	10.0	27.5	21.0	1.3	6.54	141
	BLUE	7.0	1.0	0.2114	0.012	17.6	88.0	1.1

Table 3.2 Summary of various approaches used in the purification of GDH.

The abbreviations used are as follows: CE = Cell extract, and the various chromatography steps employed were: RG = Matrex Gel Red A, GF = Gel filtration, H-SP = HiTrapSP (cation exchange) and BLUE = F3GA Cibacron Blue.

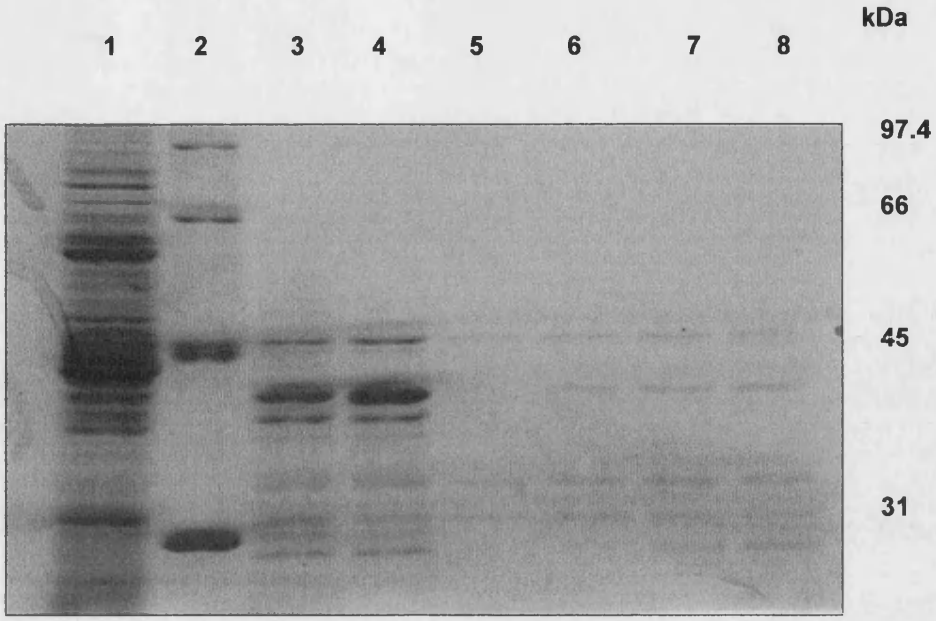


Fig. 3.1 SDS-PAGE of cell extract and the samples purified using approach 1.

(1) Cell extract, (2) Molecular weight markers, (3) & (4) fractions from Matrex Gel Red A column, (4) - (8) fractions from Gel filtration column.

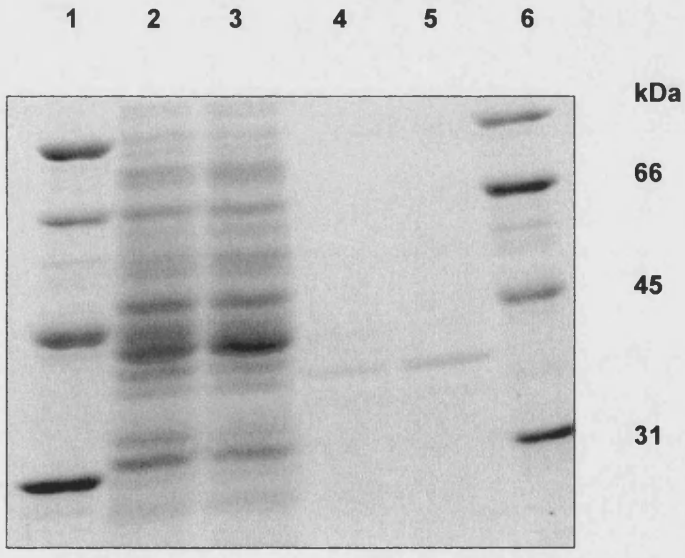


Fig. 3.2 SDS-PAGE of purified samples using approach 2.

(1) & (6) Molecular weight markers, (2) & (3) Gel filtration samples and (4) & (5) are Matrex Gel Red A samples.

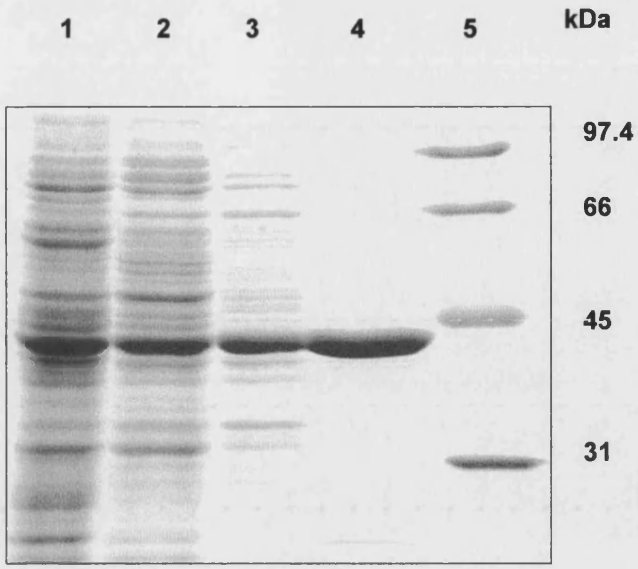


Fig. 3.3 SDS-PAGE of purified samples using approach 3.

(1) Cell extract, (2) Gel filtration sample, (3) HiTrap SP sample, (4) Matrex Gel Red A sample and (5) Molecular weight markers.

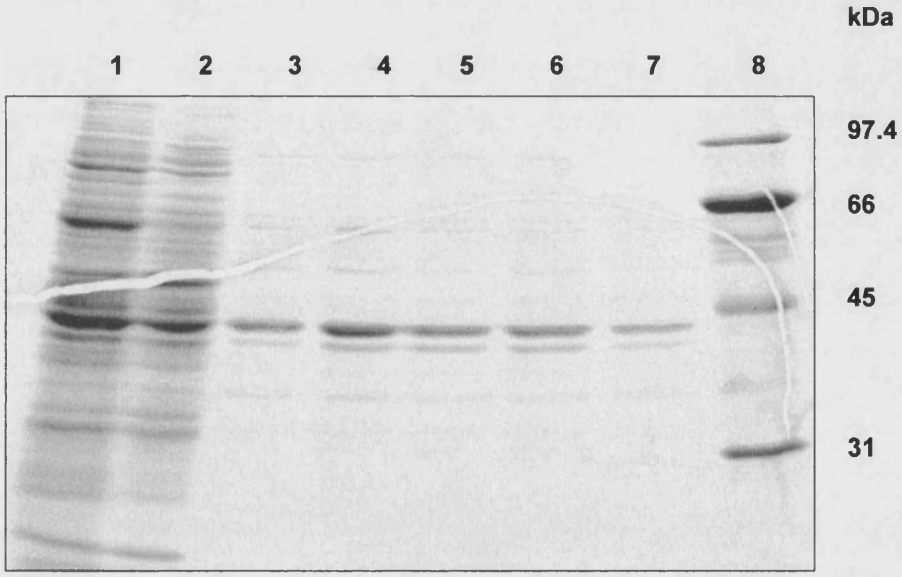


Fig. 3.4 SDS-PAGE of purified samples using approach 4.

(1) Cell extract, (2) Gel filtration sample, (3) - (7) fractions from Matrex Gel Red A column and (8) Molecular weight markers.

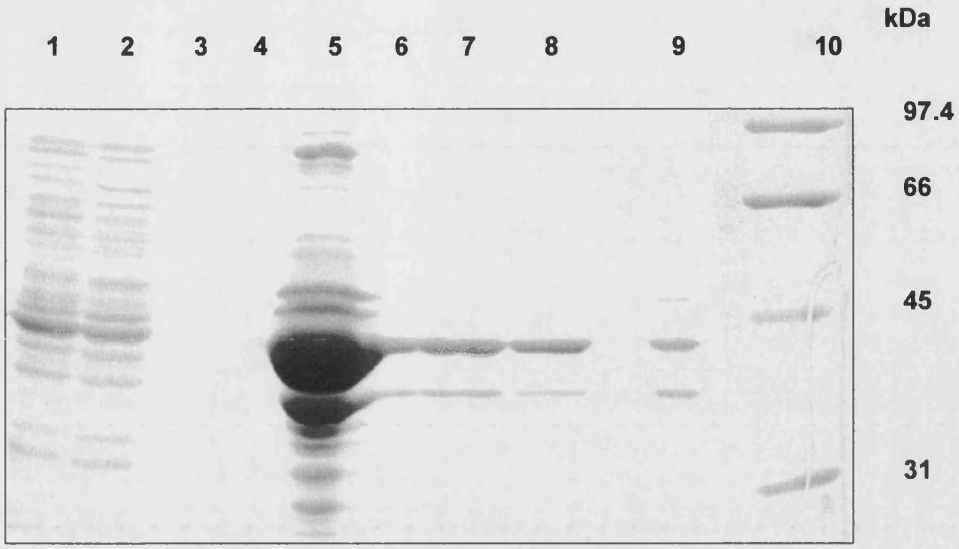


Fig. 3.5 SDS-PAGE of purified samples using approach 5.

(1) Cell extract, (2) Gel filtration sample, (3) & (4) fractions from HiTrap SP column following the affinity chromatography step, (5) - (9) fractions from Affinity Cibacron Blue column and (10) Molecular weight markers.

3.3.2 IDENTIFICATION OF THE GDH PROTEIN USING ACTIVITY STAINING AND DETERMINATION OF THE N-TERMINAL SEQUENCE

3.3.2.1 Trials 1&2 - Western blotting and Native-PAGE activity staining

Following the first successful purification (approach 1), the purified samples were transferred to PVDF membrane using the Western blot procedure as described in 3.2.2.4. Based on their migration distances, the five protein bands yielded by approach 1 (Fig. 3.1) gave the following Mr values shown in Table 3.3. With reference to the Mr of GDH from *T. acidophilum* (≈ 40 kDa) (Smith *et al.*, 1989) and from *S. solfataricus* (≈ 31 kDa) (Giardina *et al.*, 1986), it was decided that the protein bands nearest the 45 kDa and the 31 kDa protein markers, respectively, would be subjected to N-terminal sequencing analysis. The resulting sequences were as follows:

Band 1 (48 kDa)

MKIGLIGLGIMGYRIAANLAKANKLNLVYDRTQE?IE(R)

Band 4 (29 kDa)

?LGIAFVG(S)GF?A(K)F?L(R)?L

Protein	Distance migrated (cm)	Log Mr	Mr/10^3
Band 1	3.5	1.68	48
Band 2	3.95	1.62	42
Band 3	4.8	1.50	32
Band 4	5.1	1.46	29
Band 5	5.35	1.42	26

Table 3.3 Migration distances of purified proteins (protein bands 1 – 5) and their corresponding Mr values.

Calculated based on the proportional relationship of a protein-SDS complex to Log Mr.

On processing these sequences through the SWISSPROT database, it was discovered that neither of them possessed sequence identity to any of the glucose dehydrogenase sequences.

However, further approaches to purifying GDH (approaches 2 - 5) revealed that a protein band of Mr 40 - 43 kDa is a more likely candidate for the GDH and is almost purified to homogeneity using approaches 3 and 4, although it was decided that samples purified by approaches 1 (Fig. 3.1, lanes 4 - 8) and 5 (Fig 3.5, lane 5) were to be used for activity staining, as the pH employed in procedures 3 and 4 were not favourable for maintaining GDH activity. However, the result of cutting out an activity-stained band from Native-PAGE failed to ensure that only one band could be removed and subsequently subjected to SDS-PAGE for final analysis. Therefore, it was decided to use the procedure outlined in 3.2.2.6, which would allow the satisfactory separation of the proteins and thus aid in the isolation of the solitary protein band responsible for GDH activity.

3.3.2. Trial 3 - SDS-PAGE activity staining

A prominent, activity-stained band of approximate Mr of 40 - 43 kDa was cut out after SDS-PAGE of a sample that had been purified by approach 5 (Fig. 3.5, lane 5). The SDS-PAGE had been previously washed for two hours immediately following electrophoresis (as

described in 3.2.2.6) and was not subjected to staining with Coomassie. Subsequent electrophoresis of this band on another SDS-PAGE stained with Coomassie (Fig. 3.6), revealed a prominent band of Mr 40 - 43 kDa and another less conspicuous band below it of approximate Mr 40 kDa, which was carried over during the excision procedure.

This protein band (Band 3) of Mr 40 - 43 kDa was then subjected to N-terminal sequencing. The following sequence was obtained:

MKAIIVKPPNAGVQVKDVDEK??D??

A Swissprot database search of this sequence revealed the following matches:

1. A 40% identity with malate dehydrogenase (NAD) at the region of 90-97 aa.
2. A 37% identity with a hypothetical oxidoreductase in RTP 5' region at the region of 17-23 aa.
3. A 36.2% identity with adenylate kinase (EC 2.7.4.3) (ATP-AMP transphosphosphorylase) from *Schizosaccaromyces pombe* at the region of 1-5 aa.
4. A 36.2% identity with glucose dehydrogenase (EC 1.1.1.47) from *T. acidophilum* at the region of 1-7 aa.
5. A 35.2% identity with alcohol dehydrogenase (EC 1.1.1.1) (ADH-T) from *Bacillus* at the region of 1-7 aa.

Other matches included many proteins such as those involved with transcription, nuclear factors and heat shock proteins from bacteria and animalia.

A comparison table using the PILEUP facility in the GCG programme produced the following result shown Table 3.4

kDa

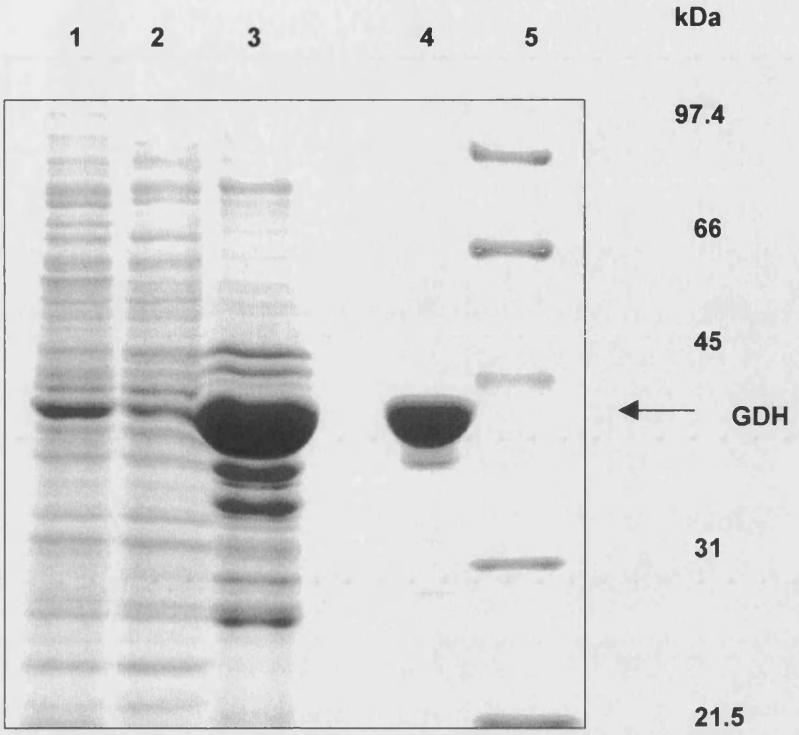


Fig. 3.6 **SDS-PAGE of activity-stained band excised from washed SDS-PAGE.**

(1) Cell extract, (2) Gel filtration sample, (3) original Affinity Cibacron Blue sample, (4) Activity-stained band from washed SDS-PAGE of Affinity Cibacron Blue sample and (5) Molecular weight markers.

S. solfataricus GDH (1-24 aa)	.	.	.	M	K	A	I	I	V	K	P	.	P	N	A	G	V	Q	V	K	D	V	D	E	K	X	X	D	.	.
S. solfataricus ADH (1-24 aa)	.	M	R	A	V	R	L	V	E	I	G	K	P	L	.	.	S	L	Q	E	I	.	G	V	P	K	P			
T. tenax GDH (1-24 aa)	.	M	R	A	V	T	.	V	T	P	G	V	P	E	.	.	S	L	R	L	R	.	E	V	P	E	P			
T. acidophilum GDH (1-30 aa)	M	T	E	Q	K	A	I	V	T	D	A	.	P	K	G	G	V	K	Y	T	T	I	D	M	P	E	P	E	H	Y
H. mediteranii GDH (1-27 aa)	.	.	.	M	K	A	I	A	V	K	R	G	E	D	R	P	V	V	I	E	K	P	R		P	E	P	E	S	G

Table 3.4 Alignment of the MDR dehydrogenase N-terminal sequences from Archaeal species.

The *S. solfataricus* GDH N-terminal sequence represents the N-terminal sequence obtained from the protein purified in this work.

page 49 ?

3.4 DISCUSSION

3.4.1 GDH PURIFICATION

20 mM $MgCl_2$ was added to all gel buffers employed in the purification of GDH, as it has been reported that the enzyme requires the presence of bivalent cations for stability (Giardina *et al.*, 1986).The approach that gave the highest level of purified GDH from *S. solfataricus* was approach 3 (Fig 3.1). It utilised the following series of steps : Gel Filtration, Cation Exchange and Matrex Gel Red A with buffers at pH 5.5. This pH was necessary in order to facilitate good binding of GDH to the column matrix of the cation exchange and the Red gel A columns (Garg *et al.*, 1996). However despite providing a good purification, the GDH enzyme did not appear to be stable at such a low pH and all the activity was lost after 24 h. Also, the long duration of this purification process and subsequent dialysis to a more suitable pH (for example pH 8) after the last step did not prevent the permanent loss of the majority of the activity. However, although it initially appeared to be the low pH that was responsible for loss of activity, knowledge of the internal pH of *S. solfataricus* reveals it to be about pH 5.5 (Lübben and Schäfer, 1989), possibly arguing against this idea. Therefore, it could be that the presence of high salt concentrations used to elute the enzyme (approximately 1-1.5 M NaCl), or the high salt in conjunction with the low pH, are responsible for the loss of activity

Another more suitable approach to purification was found, which would maintain GDH nearer to its favoured pH of 8-9 (Giardina *et al.*, 1986) and which used an Affi-Gel Blue column as the last purification step. This was carried out using buffers at pH 7 and the elution step involved the use of $NADP^+$ and xylose. Therefore, the requirement for a more favourable pH and the use of a substrate specific elution, in preference to high-salt elution, made this an ideal step to employ for further purification trials. The efficiency of this approach was confirmed by the isolation of the GDH protein and the obtaining of a N-terminal sequence. However, there are in fact some disadvantages to this purification method. Firstly, the purification fold was not of the magnitude achieved by previous approaches and it would benefit from an additional step, for instance, anion exchange. Also secondly, the use of substrates, in particular that of coenzyme, makes this method very costly, especially if the process required further scaling up.

3.4 DISCUSSION

3.4.1 GDH PURIFICATION

20 mM $MgCl_2$ was added to all gel buffers employed in the purification of GDH, as it has been reported that the enzyme requires the presence of bivalent cations for stability (Giardina *et al.*, 1986).The approach that gave the highest level of purified GDH from *S. solfataricus* was approach 3 (Fig 3.1). It utilised the following series of steps : Gel Filtration, Cation Exchange and Matrex Gel Red A with buffers at pH 5.5. This pH was necessary in order to facilitate good binding of GDH to the column matrix of the cation exchange and the Red gel A columns (Garg *et al.*, 1996). However despite providing a good purification, the GDH enzyme did not appear to be stable at such a low pH and all the activity was lost after 24 h. Also, the long duration of this purification process and subsequent dialysis to a more suitable pH (for example pH 8) after the last step did not prevent the permanent loss of the majority of the activity. However, although it initially appeared to be the low pH that was responsible for loss of activity, knowledge of the internal pH of *S. solfataricus* reveals it to be about pH 5.5 (Lübben and Schäfer, 1989), possibly arguing against this idea. Therefore, it could be that the presence of high salt concentrations used to elute the enzyme (approximately 1-1.5 M NaCl), or the high salt in conjunction with the low pH, are responsible for the loss of activity

Another more suitable approach to purification was found, which would maintain GDH nearer to its favoured pH of 8-9 (Giardina *et al.*, 1986) and which used an Affi-Gel Blue column as the last purification step. This was carried out using buffers at pH 7 and the elution step involved the use of $NADP^+$ and xylose. Therefore, the requirement for a more favourable pH and the use of a substrate specific elution, in preference to high-salt elution, made this an ideal step to employ for further purification trials. The efficiency of this approach was confirmed by the isolation of the GDH protein and the obtaining of a N-terminal sequence. However, there are in fact some disadvantages to this purification method. Firstly, the purification fold was not of the magnitude achieved by previous approaches and it would benefit from an additional step, for instance, anion exchange. Also secondly, the use of substrates, in particular that of coenzyme, makes this method very costly, especially if the process required further scaling up.

3.4.2 ACTIVITY STAINING

The Native-PAGE activity stain successfully managed to reduce the number of protein bands visible on SDS-PAGE after the stained band was cut out of the Native-PAGE. However, the poor resolution obtained with Native-PAGE inevitably leads to the transfer of neighbouring proteins, which explains the appearance of more than one band when this sample was run on a subsequent SDS-PAGE. However, this problem was rectified by performing an SDS-PAGE in the normal way and then carrying out a 2 h washing step to remove the SDS (described in 3.2.2.6). Following this procedure, the activity stain was carried out (section 3.2.2.5) and the resulting coloured band removed and run on another SDS-PAGE (Fig. 3.5). This method takes advantage of the good resolution provided by SDS-PAGE and the successful removal of SDS permitting the renaturing of the GDH protein. This would confirm or suggest that *S. solfataricus* GDH like other nicotinamide-nucleotide-dependent GDHs, consists of homogeneous subunits, due to the fact that the subunits would all have to migrate in synchrony with one another to produce only one activity-stained band. As studies on the GDH from *Bacillus megaterium* showed that dissociation of the tetrameric enzyme yields inactive monomers (Pauly and Pfleiderer, 1977), therefore the subunits must be in close proximity with one another in order to reassociate and become active again. Also the procurement of one N-terminal sequence from this protein band confirms its homogeneous state.

3.5.3 N-TERMINAL SEQUENCING

The SDS-PAGE produced from purification approach 1 narrowed down the number of candidates (protein bands) for GDH to five. The two particular bands that were then sequenced were chosen for their relative migration distances and due to the fact that they co-migrated with the 45 kDa and 30 kDa marker proteins respectively. These Mr values are relevant as many glucose dehydrogenases, including that of *T. acidophilum,* are approximately 40 - 45 kDa and the GDH from *S.* solfataricus isolated by Giardina *et al.* (1986) is 30 kDa. Although protein Bands 1 and 4 were not found to contain GDH and in fact the fact the band below this, at 42 kDa, should have been sequenced for the previously mentioned reasons. A database search through the Swissprot database showed that in fact the Band 1 protein most closely resembles 6-phosphogluconate dehydrogenase (EC 1.1.1.44) from many bacteria and mammals. It also had no sequence similarity with *T. acidophilum* GDH and was therefore eliminated as a

candidate for GDH. The Band 4 protein has a Mr of 29 kDa, which is smaller than the value of *T. acidophilum* GDH but is very similar to the molecular weight of 31 kDa proposed for the *S. solfataricus* GDH by Giardina *et al.* (1986). However, it was only found to match with an *E. coli* hypothetical 41.1 kDa protein (EC 1.18.1.-). Also the fact that its Mr is a lot lower than that of the majority of glucose dehydrogenases prompted further searching. In view of the purification of a GDH from *Sulfolobus solfataricus* by Giardina *et al.* (1986), that had a Mr of 30 kDa and also taking into account the purification carried out in this report, it is probable that *Sulfolobus* may contain more than one glucose dehydrogenase - with different specificities for glucose. However one of the purification methods (Approach 5) employs an elution step containing the substrates xylose and coenzyme as used in Giardina *et al.* (1986) but this together with the other approaches consistently yielded only one protein of approximately 40 kDa. It is additionally important to note that activity staining of a sample of the cell extract on non-denaturing-PAGE also only yields one activity-stained band (results not shown).

The Band 3 protein has a molecular weight of approximately 40 kDa and correlates well with that of the *T. acidophilum* GDH. A database search using its N-terminal sequence also revealed a match with the *T. acidophilum* GDH giving a 36.2% identity. This correlates well with the alignment of this protein sequence against the N-terminal sequences of *T. acidophilum* and other nicotinamide-nucleotide-dependent-dehydrogenases, namely *T. acidophilum* GDH, *T. tenax* GDH, *S. solfataricus* ADH and *H. mediteranni* GDH which showed several amino acid matches (Table 3.4). Thus confirming the identity of this protein as that of the GDH enzyme. These particular dehydrogenases and their species were chosen due to the fact that they are all thermophilic archaeal species (*H. mediteranni* is a halophile, but its GDH also possess thermophilic properties (Bonete et al., 1996)) and more importantly *T. tenax* was chosen due to its proximity to *Sulfolobus* in the rRNA phylogenetic tree. The ADH was also included due to its similarity in relative molecular mass to other nicotinamide-requiring GDHs and it also groups together with *T. acidophilum* GDH within the MDR family (Jörnvall et al., 1987; see chapter 1).

4 Chapter Four

Cloning, sequencing and analysis of the Glucose dehydrogenase gene

4.1 INTRODUCTION

In order to isolate the GDH enzyme and obtain it in large enough quantities for characterisation and crystallisation studies, it was necessary to isolate the GDH gene and clone it into a suitable expression vector. At the same time, the sequencing of the cloned gene would provide further information about its evolutionary origin in terms of its grouping with other pyrimidine-nucleotide-dependent alcohol/polyol/sugar (medium-chain, former long-chain, frequently zinc-containing) dehydrogenases (Jörnvall *et* al., 1984; Jörnvall *et* al., 1987; Persson *et al.*, 1991).

4.2 MATERIALS AND METHODS

4.2.1 MATERIALS

All reagents were of analytical grade or of the finest grade commercially available.

4.2.1.1 Organisms

E. coli TG1 [δ(*lac-pro*)*thi sup*E [Res$^-$Mod$^-$(k)]F(*tra*D36*pro*A$^+$B$^+$ *lac*I^{q}ZδMI5] was originally from A. Bankier, MRC Cambridge, U.K. *E. coli* XL Blue MRA (P2) cells (*E. coli*, Δ(*mcr*A) 183 Δ(*mcr*CB-*hsd*SMR-*mrr*) 173 *end*A1 *sup*E44 *thi*-1 *gyr*A96 *rel*A1 *lac* (P2 lysogen) were supplied by Strategene, La Jolla CA, USA. JM109 high-competent cells (*Eschericia coli*, e14$^-$(McrA$^-$) *rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 (rκ$^-$ rκ$^-$) *sup*E44 *rel*A1 Δ(lac-proAB) [F' *tra*D36 *pro*AB *lac*I^{q}ZΔM15]) were supplied by Promega, Madison, WI, U.S.A. The λ Phage library of the *Sulfolobus solfataricus* genome was supplied by Dr Helen Connaris , University of Bath.

4.2.1.2 Vectors

Plasmid pMEX8 was originally supplied by Medac GmBH and was used in the construct pTaGDH8 (Bright et al., 1991). pGEM®-T vector system was obtained from Promega,

Madison, WI, U.S.A. pREC7/NdeI was supplied by Dr Linda Kurtz, Washington University, Missouri, U.S.A. pET-3a vector system was obtained from Novagen, Abingdon, UK.

4.2.1.3 Enzymes

Restriction endonucleases and reaction buffers were obtained from New England Biolabs Inc., MA, U.S.A. Pwo/ Taq Mix and reaction buffer were supplied by Hybaid Ltd., Teddington, Middlesex, U.K. T4 DNA ligase and Klenow enzyme were supplied by Boehringer-Mannheim, Mannheim, Germany.

4.2.1.4 Reagents

Tryptone for LB culture media was purchased from Amersham, Bury, U.K. Yeast extract and bactoagar were from Difco, Michigan, USA. 1 Kb DNA Ladder and Low DNA Mass Ladder™ molecular mass markers for agarose gels and Tris were from GIBCO BRL Life Technologies, New York, U.S.A. SeaKem LE agarose was obtained from FMC BioProducts, Rockland, ME, U.S.A. and Low Melting Point Agarose was from Schwarz/Mann Biotech, Cleveland, Ohio, U.S.A. Designed primers were supplied by Perkin-Elmer Applied Biosystems, Warrington, Cheshire, UK. Random Primed and 3' End Labelling Kits were from Boehringer, Mannheim, Germany. Hybond N^+, Hyperfilm™ MP (High performance autoradiograph film) and [α-32P]dCTP (9.25 Mbq; 250μCi) were from Amersham International, Little Chalfont, Buckinghamshire, U.K. QIAGEN Plasmid Midi prep kit and QIAGEN Lambda Maxi Kit (10) were from QIAGEN GmBH, Hilden Germany. Wizard™ Plus Minipreps System and WIZARD® Lambda Preps DNA Purification System were obtained from Promega, Madison, WI, U.S.A. Geneclean®, Geneclean®SPIN and MERmaid® DNA extraction kits was from BIO-101, Inc., Vista, CA, U.S.A. G-50 Sephadex matrix, and M13 sequencing and reverse primers were supplied by Pharmacia, Uppsala, Sweden. IPTG and X-gal were supplied by Calbiochem, La Jolla, CA, U.S.A and Immobilon™-P Transfer membrane was from Millipore Corp., Bedford, U.K.

4.2.1.5 Equipment

The Compact Line - Hybridisation Oven OV4, was purchased from BIOMETRA Ltd., Maidstone, Kent, U.K. The Spectrophotometers, Lamda 3B UV/Vis and Lambda Bio 230V UV/Vis/NIR, and the Cetus DNA Thermal Cycler were supplied by Perkin-Elmer Corporation, Norwalk, Connecticut, U.S.A. Milli-Q water was from the Milli-Q™ reagent-grade water system

from Millipore Corp., Bedford, MA, U.S.A. ABI PRISM™ 377 DNA sequencer was from Applied Biosystem, Fostericity, CA, U.S.A.

4.2.2 METHODS

4.2.2.1 Storage and culture of *Escherichia coli* TG1

E.coli TG1 cultures containing the plasmid pTaGDH8 were stored in 20% (v/v) sterile glycerol at -20°C and cultured when required by shaking and incubation at 37°C in LB broth containing 0.1% (w/v) ampicillin.

4.2.2.2 Storage and culture of *Escherichia coli* P2 Lysogen

E.coli P2 cultures were stored on minimal agar at 4°C and sub-cultured every 4 - 8 weeks; cells were grown by incubation at 37°C in LB broth, 0.2% (w/v) maltose and 10 mM $MgSO_4$ with shaking or on LB agar plates.

4.2.2.3 Storage and culture of *Escherichia coli* JM109

E. coli JM109 cultures were stored on minimal agar at 4°C and sub-cultured every 4-8 weeks; cells were grown by incubation at 37°C in LB broth with shaking or on LB agar plates.

4.2.2.4 Determination of DNA concentration and purity

The concentration of DNA in aqueous solutions was determined spectrophotometrically by measurement of absorbance at 260 nm (Sambrook *et al.*, 1989), using the following approximations:

1 absorbance unit ≡ 50 μg ds DNA/ml

1 absorbance unit ≡ 30 μg ss DNA/ml

DNA purity was assessed spectrophotometrically by measurement of absorbance at 260 nm and 280 nm and calculation of the A_{260}/A_{280} ratio; a value ≥1.8 for ds DNA and ≥1.6 for ss DNA were the criteria for purity.

4.2.2.5 Agarose gel electrophoresis

Examination of solutions containing DNA by agarose gel electrophoresis allows the separation of linear, duplex DNA of differing sizes between 0.8 - 10 Kb (Sambrook *et* al., 1989). Agarose was dissolved in boiling TAE buffer (40 mM Tris acetate, 1 mM EDTA) and allowed to cool to <50°C; ethidium bromide was added to a final concentration of 400 μg/l and the solution

was used to pour slab gels with the addition of a well-forming comb. Standard agarose gels were prepared at a concentration of 1.0% (w/v) and gels containing Low Melting Point agarose were prepared at a concentration of 0.7% (w/v).

Once set, the well-forming comb was removed and DNA in 6x sample buffer (40% (w/v) sucrose, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol), at a concentration ≤ 200 ng DNA/ml, was loaded onto slab gels and separated by electrophoresis in 1x TAE at a constant voltage of 70 V until the xylene cyanol dye front was 1 cm from the bottom edge of the gel. The DNA bands were then visualised using a UV transilluminator. A 1 Kb Ladder™ was used as the molecular weight marker.

4.2.2.6 Isolation of DNA fragments from agarose gels

DNA fragments were first isolated from the (low melting point) agarose gel as gel slices. Further steps were carried out according to the manufacturer's instructions. Geneclean® and MERmaid® methods use silica-based matrices called Glassmilk® and Glassfog®, respectively, that have high affinity for DNA.

4.2.2.7 Plasmid and Lambda DNA "Mini-Maxipreps"

DNA plasmid "Midipreps" were performed using QIAGEN Plasmid or Wizard™ Plus plasmid preparation kits according to the manufacturer's instructions. A MAXI λ DNA preparation kit (QIAGEN Lamda Maxi Kit (10) DNA or WIZARD® Lambda Preps) was used to prepare DNA from cultured λ bacteriophage as per manufacturer's instructions.

4.2.2.8 Restriction of DNA

Digestion of DNA using endonucleases was carried out as specified in the manufacturer's instructions. DNA was incubated with restriction enzyme(s), at a final concentration of 2U enzyme(s)/μg DNA, at 37°C in an appropriate restriction buffer (supplied with the enzyme as a 10x concentrate). Incubation at 37°C varied from 1 h to overnight depending on the "star" activity of the endonuclease (the ability to cleave similar but not identical sequences under extreme conditions e.g. prolonged incubation). Some restriction enzymes also required the presence of 100 μg/ml of BSA for activity.

4.2.2.9 Amplification of DNA - Polymerase chain reaction (PCR)

Reactions generally contained approximately 100 ng of template DNA, 100 μmol of

each primer, and 10 mM each of dATP, dCTP, dGTP and dTTP, in a final volume of 50 µl. The reactions were carried out using the Pwo/Taq Mix enzyme and were buffered with 10x Reaction buffer-Incomplete (500 mM Tris/HCl, pH 9.1, 150 mM $(NH_4)_2SO_4$, 20% (v/v) DMSO, 1% (v/v) Tween-20) plus 1.5 mM $MgCl_2$. The reaction mixture was overlaid with 50 µl of mineral oil to prevent evaporation. Reactions were incubated for 5 min at 94°C and 85°C for 2 min prior to initiation of the PCR, in order to provide a "hotstart" on adding the enzyme. Amplification was carried out in cycles as follows: denaturation at 94°C for 1 min, annealing at 55°C for 2 min and extension at 72°C for 2 min. On finishing the 30 cycles, a further 10 min (72°C) extension cycle was carried out in order to complete all the amplification reactions. A 10 µl sample was then visualised on a 2% (w/v) agarose gel.

4.2.2.10 Screening procedures

4.2.2.10.1 Preparation of top-agar plates for screening gDNA λ library

A single colony of XL1 Blue MRA cells (P2) was used to inoculate 20 ml of LB supplemented with 10 mM $MgSO_4$ and 0.2% (w/v) maltose in a sterile 50 ml conical tube, and the suspension was incubated in a shaking incubator at 37°C until an OD_{600} of 1.0 was reached. The P2 cells were immediately pelleted at 1,500 x g for 10 min. The pellet was resuspended in ice-cold 10 mM $MgSO_4$ to give a final OD_{600} of 0.5. The cells were stored on ice until needed (it is possible to store them for up to 24 h at 4°C). 1 - 10 µl (serial dilutions of 10^0,$10^{-1}$$10^{-2}$, 10^{-3} and 10^{-4}) of λ library stock was used to infect 200 µl of P2 cells per plate to be screened. The mixture, contained in a sterile 15 ml conical tube, was incubated at 37°C for 20 min with occasional agitation. 3 ml of molten top-agar (LB broth supplemented with 0.7% (w/v) agarose) at 48°C was added to each 15 ml conical tube prior to mixing by inversion and immediately poured on to pre-warmed agar plates (37°). On setting (approximately 30 s), the plates were then incubated at 37°C overnight in a plate incubator.

4.2.2.10.2 Transfer of λ clone DNA from top-agar plates to nylon membranes

Each plate and membrane were numbered (using a soft pencil) in order to facilitate identification at a later date. The membrane was gently lowered, centre first, on to the surface of the top-agar plate. Care was taken to ensure that no air bubbles were trapped underneath it. Once in position, the membrane and agar were stabbed using a hyperdermic needle. These

stab positions were also marked on the outside of the agar plate using a permanent ink marker. This would ensure that the plate and the autoradiograph could be re-aligned later on. After 1 - 2 min, the membrane was carefully removed using a pair of forceps and left DNA side up on paper towels to air dry for 15 min.

4.2.2.10.3 Denaturation of DNA bonded to nylon membranes

Once the DNA was transferred to the nylon membranes, it was denatured in order to fix it to the membranes and allow the hybridisation of probes. Three trays containing 3MM Whatman paper cut larger than the membranes were saturated with (1) Denaturation solution, 0.2 M NaOH, 1.5 M NaCl, (2) Neutralisation solution, 0.4 M Tris-HCl, pH 7.6, and (3) 2x SSC. Using forceps, each membrane (DNA side up) was incubated in trays 1 - 3, for 7, 7 and 2 min at room temperature respectively. Care was taken whilst transferring from one tray to the next to minimise the transfer of solutions. After the final incubation, the membranes were left DNA side up on 3MM Whatman paper to air dry for 5 min before a final fixing step of incubation in an oven at 80°C for 1 - 2 h.

4.2.2.11 Southern transfer

DNA fragments in agarose gels were transferred and bound to Hybond-N hybridisation membrane according to the manufacturer's instructions (capillary blot procedure of Southern, 1975); the membrane was washed in 6x SSC, fixed at 80°C for 1 - 2 h, sealed and stored at -20°C until required for hybridisation.

4.2.2.12 ^{32}P-labelling of probe

The method of "random primed" labelling was adopted for this procedure using the Random Primed Labelling Kit. Approximately 25 ng of denatured DNA (probe) was combined with deoxynucleoside triphosphates, reaction mixture, $\cong$ 50 μ Ci [α^{32}P] dCTP, sterile redistilled water and Klenow enzyme (for exact concentrations refer to manufacturer's instructions) for 30 min at 37°C. The reaction was stopped with 0.2 M EDTA (pH 8.0) and the non-incorporated deoxyribonucleoside triphosphates were removed by chromatography through G50 sephadex (Fine).

4.2.2.13 Hybridisation of the labelled probe to the membrane

The dry membrane filters were placed in Biometra flasks with approximately 5 ml of

prehybridisation buffer (5% (w/v) Dextran sulphate, 0.5% (w/v) Blocking agent, 0.1% (w/v) SDS in 5x SSC and 50μl of denatured salmon sperm DNA per filter). The flasks were incubated for 2 h initially at 55°C and subsequently at 45°C in a rotating hybridisation oven. After the pre-hybridisation step, 5 μl of the ^{32}P-labelled probe was added and the flasks were incubated shaking overnight at 45°C.

4.2.2.14 Washing and autoradiography of the labelled membrane

Originally the membranes were washed as follows (Type 1) :

1. Twice in 2 x SSC, 0.1% SDS, for 15 min at 55°C.
2. Twice in 1 x SSC, 0.1% SDS, for 15 min at 68°C.

Subsequent washes (Type 2) were as shown:

1. Twice in 1 x SSC, 0.5% SDS, for 15 min at 45°C
2. Twice in 0.5 x SSC, 0.5% SDS for 15 min at 45°C.

The membrane were sealed in plastic wrap film, over-laid with pre-flashed X-ray film, placed in an autoradiography cassette lined with intensifying screens and incubated at -70°C for 24 - 48 h and developed using a Hyperprocessor. The intensity of the image developed on this film would determine whether it would be necessary to develop another film for a longer period of time.

4.2.2.15 Stripping membranes

The membranes were incubated for 2h at 75°C in :

1 mM Tris/HCl, pH 8.0

1 mM EDTA

0.1% Denhardts Reagent

The membranes were then dried and placed in plastic wrap film until required for pre-hybridisation.

4.2.2.16 Isolation and storage of λ clone

Once a positive clone was identified from the autoradiographs and the corresponding plaque found, a plaque plug was taken. A sterile Pasteur pipette was lowered on to a specific plaque so that a "plug" of agar containing that single plaque is removed from the plate. The plug was then expelled from the Pasteur pipette into a sterile 1.6 ml microfuge tube containing 1 ml

of SM buffer (0.58% (w/v) NaCl, 0.2% (w/v) $MgSO_4 \cdot 7H_2O$, 50% (v/v) 1 M Tris/HCl, pH 7.5, 2% (w/v) Gelatin solution) and a single drop of choloroform. The tube was then set on a vertical rotating platform and gently inverted at 4°C for 3 - 4 h. It is possible to store the resulting suspension indefinitely and it can then be used to culture a single λ clone or to perform a second screen.

4.2.2.17 Preparation of phage lysate

A single P2 colony was used to inoculate 100 ml of NZCYM medium (1% (w/v) NZ amine, 0.5% (w/v) NaCl, 0.5% (w/v) Bacto-yeast extract, 0.2% (w/v) $MgSO_4 \cdot 7H_2O$, 0.1% (w/v) Casamino acids, pH 7.0) in a 500 ml conical flask and incubated overnight at 37°C with vigorous shaking (~300 rpm). The following morning, a 5 ml aliquot of the overnight culture was used to inoculate 500 ml of fresh NZCYM medium and grown to an OD_{600} of 1.0, which corresponds approximately to 8×10^8 cells/ml. Aliquots containing 10^{10} cells were removed for each λ clone to be amplified and the cells were pelleted in 165 ml conical tubes by centrifugation at 1,500 x g for 10 min at room temperature. Each pellet was resuspended in 3 ml of SM buffer and 10 - 50 µl of plaque plug suspension was added to each aliquot and incubated at 37°C with occasional agitation. Each preparation was then added to 500 ml of pre-warmed NZCYM medium (37°C) in 2 L conical flasks. The flasks were then incubated at 37°C with vigorous shaking overnight. The following day, 10 ml of choloroform was added to the lysed cultures and incubated for a further 10 min. The λ clone DNA was then prepared as described in section 4.2.2.7.

4.2.2.18 Preparation of gDNA from *S. solfataricus*

A 0.1g pellet of *S. solfataricus* cell paste was resuspended in 567 µl TE buffer by repeated pipetting in a microfuge tube. SDS and proteinase K (20 mg/ml) were added to give a final concentration of 100 µg proteinase K /ml in 0.5 % (w/v) SDS. After through mixing and a 1 h incubation at 37°C, NaCl to a final concentration of 0.7 M was added and mixed again. 80 µl of CTAB/NaCl solution were then added and, after mixing thoroughly, it was incubated again at 65°C for 10 min, after which an approximately equal volume of chloroform/isoamyl alcohol was added and tube and contents were transferred to a microcentrifuge for a 5 min spin. The upper (aqueous) phase was then removed and transferred to another sterile microfuge tube. An equal

volume of phenol/chloroform/isoamyl alcohol was added and it underwent another centrifuge step in a microcentrifuge. The supernatant was transferred to another sterile microfuge tube and supplemented with 0.6 volumes of isopropanol. The tube and its contents were then shaken back and forth repeatedly until the appearance of stringy white DNA was observed. The pellet of DNA was then transferred to another tube containing 70 % ethanol by hooking it onto the end of a heat-sealed micropipette. After washing the DNA in 70 % ethanol, it was re-pelleted in a microcentrifuge air dried for 5 min and finally redissolved in 100 µl TE buffer.

4.2.2.19 DNA sequencing

Automated DNA sequencing was performed in house. The DNA sequencing is based on the method of Sanger *et al.* (1974), where the extension of a complementary strand is terminated with the insertion of dideoxy nucleotide (ddNTP). As each ddNTP is specifically labelled with one of four fluorescent dyes, it was not necessary to perform 4 separate reactions. The products were then analysed on a polyacrylamide gel using an ABI PRISM™ 377 DNA sequencer.

4.2.2.20 DNA ligation

Ligation of cohesive ends was carried out using 1 unit of T4 DNA ligase (at a concentration of 1 unit/µl) in the buffer supplied with the enzyme. A ratio from 1 - 3 : 1 of vector with respect to insert was used in a total volume of 10 µl. The reaction was incubated overnight at 15°C or at room temperature for 3 h. The whole reaction was used to transform competent *E. coli* cells.

4.2.2.21 Preparation of competent cells

0.5 ml of an overnight culture were grown in 100 ml sterile LB broth and incubated in a shaking incubator until it achieved an OD_{600} of 0.4 - 0.6. It was then allowed to stand on ice for 30 min before being centrifuged for 5 min, at 4°C, at 5000 x g in a bench-top microcentrifuge. The cell pellet was then resuspended in 50 ml of ice-cold 50 mM $CaCl_2$ and left to stand on ice for a further 60 min. Following this, it was centrifuged again for 5 min, at 4°C, at 5000 x g. The cell pellet was then resuspended in 10 ml of 50 mM $CaCl_2$ – 20% (v/v) glycerol and stored at -70°C in aliquots of 200µl.

4.2.2.22 Transformation

Each sample to be transformed required 50 µl of JM109 high competence cells. The cells were slowly thawed by placing on ice for 5 min. Once thawed, 1 -10 µl of a ligation reaction or control DNA was added to the cells in a sterile 1.6 ml microfuge tube and the contents were gently mixed using a pipette tip. The mixture was then stored on ice for 30 min, prior to heat-shock at 42°C for 45 s, after which the cells were immediately placed on ice for 2 min. Each transformation reaction was then transferred to a sterile 15 ml Falcon tube containing 950 µl of sterile LB broth and incubated in a shaking incubator (~150 rpm) for 1 h at 37°C. Following incubation, the culture was then transferred to a sterile microfuge tube and centrifuged for 10s at 13,000 rpm in a bench-top microcentrifuge. Approximately 800 µl of the supernatant was removed and the resulting cell pellet was resuspended in the remaining 200 µl and poured on to an agar-supplemented plate (details described in section 4.2.2.23) and incubated at 37°C overnight.

4.2.2.23 Selection of transformed cells

4.2.2.23.1 Antibiotic selection

200 µl of transformed cells were spread on to an LB agar plate supplemented with 100 µg ampicillin/ml and incubated overnight at 37°C in a plate incubator.

4.2.2.23.2 Blue/white selection

200 µl of transformed cells were spread on to a LB agar plate supplemented with 100 µg ampicillin/ml, 0.1 mM IPTG and 40 µg X-gal/ml.

4.3 RESULTS

The Results section is divided as follows: the first part (4.3.1) details the attempted isolation of the DNA sequence of the glucose dehydrogenase gene from *S. solfataricus* using the GDH gene from *T. acidophilum* as a probe in hybridisation experiments. The second part (4.3.2 and 4.3.3) deals with the alternative method of obtaining the glucose dehydrogenase sequence via employing a DNA probe based on the N-terminal sequence of the purified enzyme.

4.3.1 FIRST HYBRIDISATION (*T. ACIDOPHILUM* GENE APPROACH)

4.3.1.1 Restriction digest of pTaGDH8

In order to provide the GDH gene probe from *T. acidophilum*, it must first be amplified to provide sufficient amounts for the hybridisation experiments and additionally it must also be in a pure form. Therefore, the *E. coli* strain TG1 containing the pTaGDH8 plasmid, which includes the *T. acidophilum* GDH gene (Bright *et al.*, 1993) was grown in order to ampify the plasmid. pTaGDH8 was subsequently isolated and examined on an agarose gel in order to determine its purity (see Fig. 4.1). Isolation of the *T. acidophilum* GDH gene was facilitated by a restriction digest on pTaGDH8. This linearised the plasmid and also produced a 1.4 kb fragment containing the GDH gene, which is approximately 1 kb (Bright *et al.*, 1993) (see Fig. 4.2).

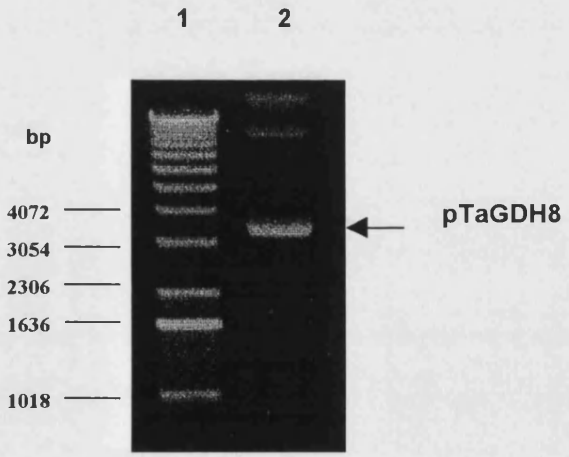


Fig. 4.1 1% agarose gel of plasmid pTaGDH8 and 1 Kb Ladder.

(1) 1 Kb Ladder and (2) pTaGDH8.

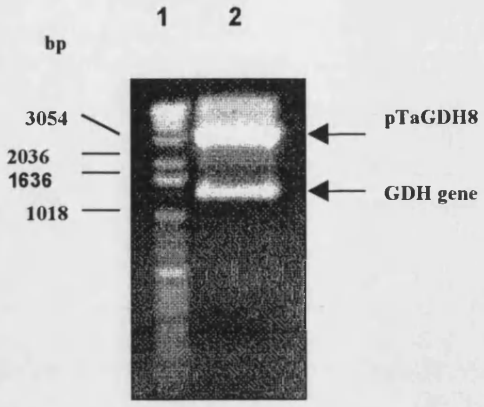


Fig. 4.2 1% agarose gel showing linearised plasmid pMEX8, 4.1 Kb and excised GDH gene, 1.4 Kb, with 1 Kb Ladder.

Following restriction of the *T. acidophilum* GDH, it was then purified using the GENECLEAN® procedure described in 4.2.2.6 and radiolabelled as described in 4.2.2.12. The resulting radiolabelled gene was then ready to undergo hybridisation (see 4.2.2.13) to the λ library containing the *S. solfataricus* genome.

4.3.1.2 Hybridisation

TRIAL 1

Following hybridisation with the radiolabelled probe (GDH gene from *T. acidophilum*) and a stringent wash programme (Type 1) as described in 4.2.2.14, it was not possible to view any evidence of hybridisation on the autoradiographic film. Therefore, the membranes were stripped and re-hybridised. They were then washed using a less stringent programme (Type 2) also described in 4.2.2.14. After 2 days incubation at -70°C, it was possible to view many hybridised plaques (radiolabelled GDH gene bound to the λ clone DNA) on all the following serially diluted plates: 10^0, 10^{-1}, 10^{-2} and 10^{-3}. (data not shown).

In order to increase the specificity, a more stringent wash was used (twice in 0.5x SSC, 0.5% SDS, at 55°C for 15 min). However this proved to be too harsh and no positive plaques were seen on developing the autoradiographic film. Therefore, it was decided that the plaques in a region of plate 10^{-2}, which gave the most intense radiation (viewed before employment of the more stringent wash) should be used for further hybridisation trials .

TRIAL 2

Following hybridisation (also with the radiolabelled *T. acidophilum* GDH gene) the

membranes were washed using the Type 2 programme. The autoradiograph produced from this trial produced many hybridised plaques after an overnight incubation. A re-wash (once 0.5 x SSC, 0.5% SDS, at 45°C for 15 min) to increase the specificity proved successful in this case and two possible candidate plaques, which should possess the GDH gene from plate $10^{-2}{}_{2}$, were isolated (A and B). (data not shown).

The plaques were removed from the plate in order to create λ stocks (4.2.2.16) and amplify by re-infecting P2 cell cultures. The phage DNA was isolated and loaded on to an agarose gel in order to evaluate the success of this procedure (Fig. 4.3).

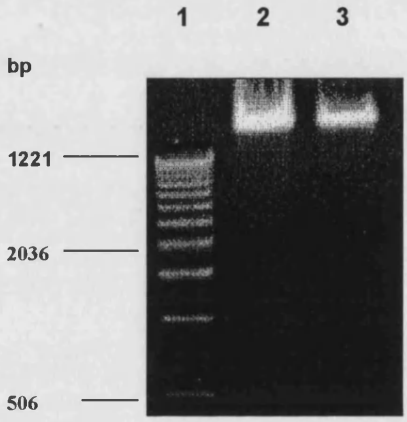


Fig. 4.3 1% Agarose gel of DNA extracted from Plaques A and B

(1) 1 Kb Ladder, (2) plaque DNA (clone A), (3) plaque DNA (clone B)

4.3.1.3 Single restriction digest of plaques A and B

The following restriction enzymes were used to cut the phage DNA :*Sal* I, *EcoR* I, *Xba* I, *Kpn* I, *Hind* III and *Bgl* I. This would provide fragments of suitable size for a Southern blot that could isolate the GDH gene on further hybridisation. However, observation of the resulting restricted fragments from these plaques (Fig. 4.4) revealed fragments that were predominantly greater that 5 Kb in size. This is too large, and fragments of approximately 3 – 5 Kb are required for accommodation into plasmid vectors and for subsequent cloning. Therefore, it was necessary to carry out double digests on the plaque DNA (Fig.s 4.5 a and b), in order to produce fragments of suitable size.

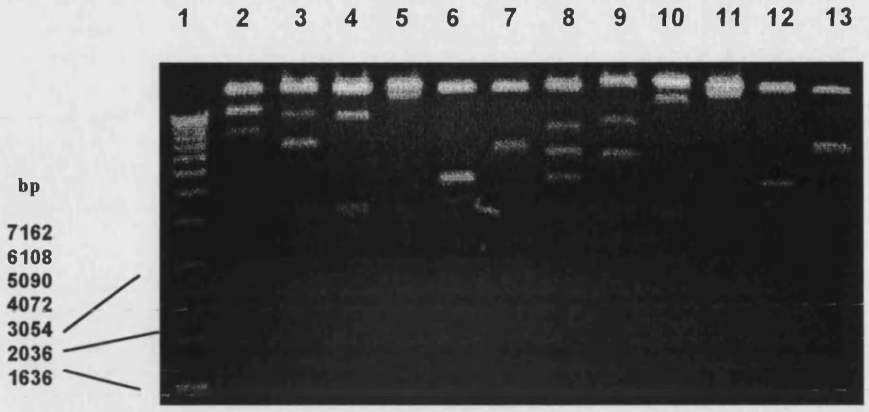


Fig. 4.4 1% Agarose gel restricted DNA from Plaque A and B.

(1) 1 Kb Ladder, (2) – (7) plaque A DNA, (8) – (13) plaque B DNA, (2) & (8) Sal I, (3) & (9) EcoRI, (4) & (10) Xba I, (5) & (11) Kpn I, (6) & (12) Hind III and (7) & (13) Bgl I.

4.3.1.4 Double restriction digest of plaque A and B DNA

After considering the suitability of various restriction buffers with the previously used enzymes, the following restriction digests were set up :

Sample Number	Restriction Endonuclease
1	1 (Sal I)
2	2 (EcoR I)
3	3 (Xba I)
4	4 (Bgl I)
5	5 (Hind III)
6	1 + 2
7	1 + 3
8	1 + 4
9	1 + 5
10	2 + 3
11	2 + 4
12	2 + 5
13	3 + 4
14	3 + 5
15	4 + 5

Table 4.1 Double restriction digests performed on DNA from Plaques A and B.

The double restriction digest of the plaque DNA produced fragments of a much more desirable size (3 – 5 kB) (Fig. 4.5 a and b). Therefore, after carrying out a Southern blot and a further hybridisation using the *T. acidophilum* GDH gene, one of these desirable fragments was found to hybridise, it could then easily be ligated into a plasmid vector.

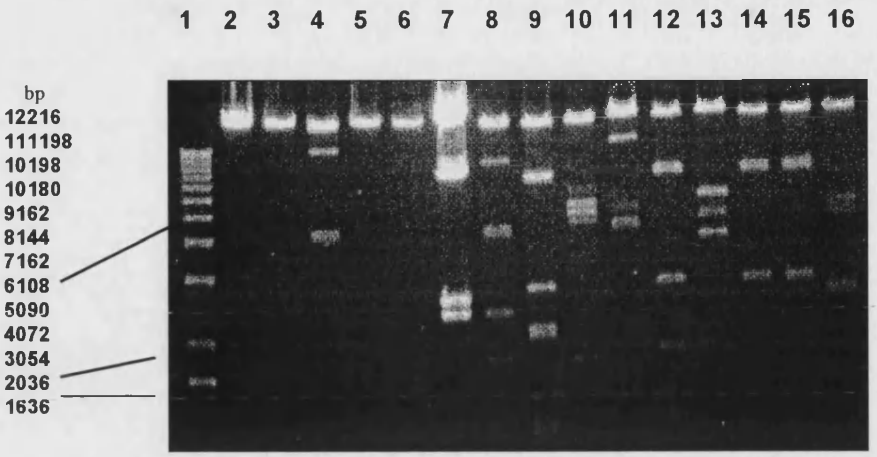


Fig. 4.5a **1% Agarose Gel showing the double restriction digest of DNA from plaque A**

(1) 1 Kb Ladder, (2) – (16) as 1 – 15 in Table 4.1.

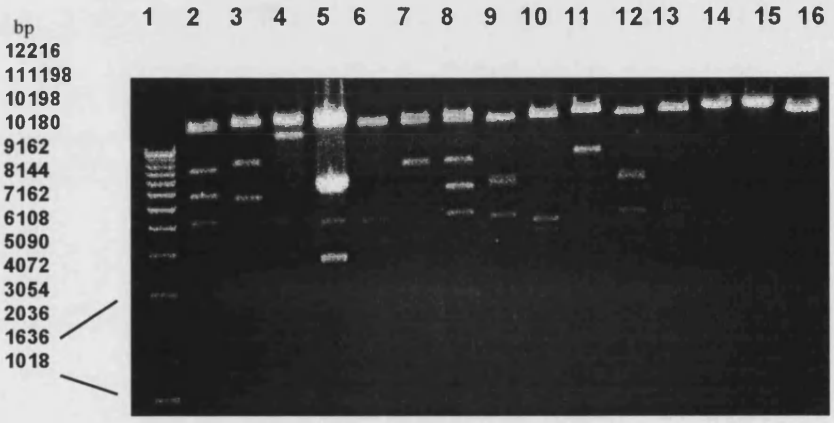


Fig. 4.5 b **1% Agarose gel of the double restriction digest of DNA from plaque B**

(1) 1 Kb Ladder, (2) – (16) as 1 – 15 in Table 4.1.

4.3.1.5 Southern blot

Following restriction digestion and agarose gel electrophoresis of the plaque DNA, a Southern Blot was performed (4.2.2.11) and the resulting bound membranes were used for hybridisation experimentation with the *T. acidophilum* GDH gene probe. The Type 2 programme of washing the membranes was employed. Unfortunately, this was unsuccessful and no evidence of hybridisation could be found on the autoradiograph.

The failure of the *T. acidophilum* GDH gene to identify the *S. solfataricus* GDH gene, led to the need to attempt a second method, namely using a DNA probe designed from the N-terminal sequence of the purified protein (Table 3.4)

4.3.2 SECOND HYBRIDISATION

4.3.2.1 Creating the 72bp probe

On obtaining the N-terminal sequence shown in Table 3.4, degenerate primers were designed to amplify by PCR the corresponding region (72 bp) of the gene using gDNA as a template (Fig. 4.6)

M K A I I V K P P N A G V Q V K D V D E K ? ? D
E P

5' ATGAARGCNATHATHGTDAAACC 3' (nina2)

5'ATGAARGCNTATHATHGTDAARCCNCCNAAYGCNGGNGTNCARGTNAARGAYGTNGAYGARAARGARCCNGAY3'
5' GAYGARAARGACCNGA 3'
3' AGNCCRAGRAARAGYAG 5'
(nina2R) 3' TCNGGYTCYTTYTCRTC 5'

Fig. 4.6 Generation of degenerate primers from the N-terminal sequence.

Amino acid and possible DNA sequences of N-terminal sequence are displayed against the forward and reverse primers, nina2 and nina2R respectively. IUB codes for the redundancies are as follows: M = A+C, R = A+G, W = A+T, S = G+C, Y =C+T, K = G+T, V = A+C+G, H = A+C+T, D = A+C+T, B = C+G+T, N = A+C+G+T.

PCR was then carried out using the degenerate primers shown in Fig. 4.6. The resulting fragment of 72 bp is shown in Fig. 4.7.

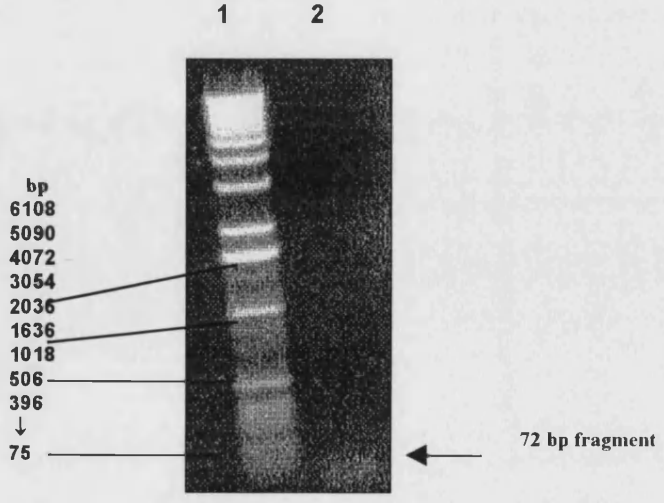


Fig. 4.7 1% agarose gel of (1) 1 KB Ladder and (2) 72 bp PCR fragment.

The 72 bp N-terminal DNA fragment was then cloned into the plasmid pGEM®-T. Further amplifications of the 72 bp probe were carried out using this recombinant vector as a template with universal primers. This provided a much greater yield and purity of amplified fragment than in PCR reactions employing the degenerate primers and genomic DNA as template. The 72 bp fragment was also sequenced to show that it did indeed correspond to the N-terminal amino acid sequence (Fig. 4.8).

M K A I I V K P P N A G V Q V K D V D E K E P E

ATGAAGGCTATCATTGTAAAACCCCCAAACGCGGGAGTGCAGGTAAAGGATGTTGATGAAAAGGAACCAGAA

Fig. 4.8 DNA sequence of the N-terminus of the *S. solfataricus* GDH gene.

Revealed by sequencing the recombinant pGEM®-T vector using universal primers. The upper line is a translation of the base sequence of the PCR product.

4.3.2.2 Hybridisation

It was decided to use the 72 bp PCR product to probe the λ library of the *S. solfataricus* genome. Following hybridisation, the nitrocellulose membranes were washed using the type 2 programme. Four positively-labelled clones were identified and isolated by secondary screening. Secondary screening involves serially diluting the isolated phage stocks and using these diluted solutions to infect more P2 *E. coli* cell cultures, in order to produce a completely pure stock of the desired clone i.e. the clone containing the GDH gene. All the secondary screens produced positive clones that hybridised to the 72 bp radiolabelled probe.

4.3.2.3 λ Maxi Prep

The four separate clones identified from the original clones were amplified by infection of P2 *E. coli* cells and the λ DNA prepared from each using the Maxi-prep method (Fig. 4.9).

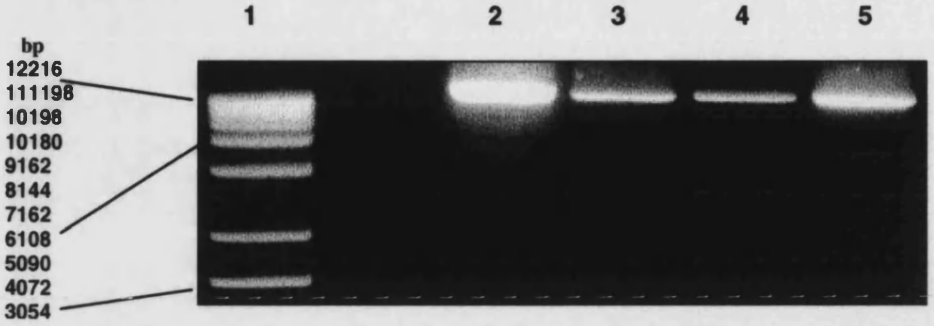


Fig. 4.9 **1% agarose gel of λ DNA obtained from positive clones**

(1) 1 Kb Ladder, (2)- (5) clones 1 - 4 respectively

4.3.2.4 Sequencing of λ clone

On the basis of the degree of DNA recovery, clone 1 was used for subsequent sequence analysis. Primers were designed from the internal region of the DNA sequence of the PCR product (Fig. 4.10). Their correspondance to the protein's N - terminal sequence meant that they were an exact match an therefore not redundant as in 4.3.2.1.

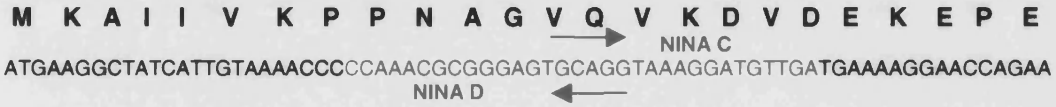


Fig. 4.10 **N - terminal sequence showing the DNA internal region used for primer design (Forward = NINA C, Reverse = NINA D).**

The following sequence information shown in Fig. 4.10 was obtained from the forward sequence reaction :

```
1       GAGATAGTTNNTGGTAAGTTAACATTGTCGACTTTACCTAAGGGAAAGGATTTCTTGGTA
60      T  L  P  K  G  K  L  T  L  S  T  L  P  K  G  K  D  F  L  V

61      TTGGGTCATGAGGCAATAGGTGTTGTGGGAAGAGTCTTATCATGGTTTTTCTCAAGGAGA
120     L  G  H  E  A  I  G  V  V  W  E  E  S  Y  H  G  F  S  Q  G

121     TCTGGTAATGCCAGTTAATANGAGAAGGATGCCGGTATTTGTANAAATTGTTTAGTTGGG
180     D  L  V  M  P  V  N  R  R  G  C  G  I  C  R  N  C  L  V  G

181     NGACCAGACTTTTGTGAACCAGGGGAATCCGGAGAAGCTGGATTCCCCANATGGTTGGGT
240     R  P  D  F  C  E  P  G  E  S  G  E  A  G  I  H  K  M  D  G  F

241     TTATGAGGGAATGGTGGNCCACGACCCTANTTATTTGGTNAAATNCCTANGGTCAATTGA
300     M  R  E  W  W  Y  D  D  P  K  Y  L  G  K  I  P  K  S  I  E
```

Fig. 4.11 Nucleotide and deduced amino acid sequence of first sequence obtained from λ clone (clone 1 DNA).

The sequence reveals the putative structural zinc-binding site (shown in red and blue) also seen in other medium-chain (former long-chain) alcohol/polyol dehydrogenases.

However subsequent sequence reactions were unsuccessful. Various strategies were undertaken to resolve this :

- removal of impurities such as host (*E. coli*) DNA and RNA, by incubating the resulting phage lysate with DNase and RNase for 1 h at 37°C prior to the removal of the phage coat.
- removal of DNases and RNases by incubation of the phage lysate with Proteinase K for 20 at 37°C prior to the removal of the phage coat.
- removal of salts from resulting λ DNA, by subjecting the sample to GENECLEAN®SPIN purification

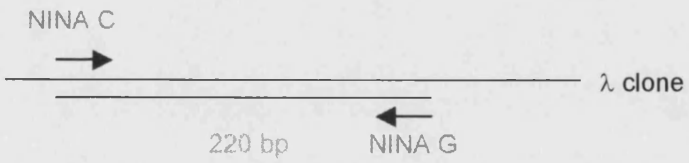
However, none of these strategies were successful, suggesting that the problem lay with the clone itself. It was then discovered that aged batches of chloroform used for storage of the bacteriophage λ (4.2.2.16) may actually inactivate it (Sambrook *et al.*, 1989). Therefore, it was necessary to repeat the hybridisation and λ clone isolation steps, this time ensuring that the resulting λ clones were stored in fresh chloroform.

Also the sequence information already obtained (Fig. 4.10) could be put to advantageous use in order to aid in the creation of a longer probe, which would greatly facilitate the re-acquisition of the λ clone containing the GDH gene. Observation of the amino acid sequences of *T. acidophilum* and other members of the MDR family, which all possess structural zinc-binding motifs, reveals that a PCR reaction with a reverse primer created at the region of the structural zinc-binding site and an N-terminal primer would produce a fragment of approximately 200 - 300 bp (Ammendola *et al.*, 1992).

4.3.3 THIRD HYBRIDISATION

4.3.3.1 220 bp Probe

On obtaining the sequence shown in Fig. 4.10, a 220 bp fragment was produced by PCR. This was produced using the primer, NINA C designed from the N - terminus as the 3' end primer and a 5' end primer which was designed from a region of the obtained sequence prior to the structural zinc binding site, NINA G.



NINA C – 5' GAGTGCAGGTAAAGGATGTTGA 3'

NINA G – 5' AACTGGCATTACCAGATCTCC 3'

Fig 4.12 PCR strategy for obtaining 220 bp probe.

NINA C was created from the central region of the N-terminal sequence (Fig. 4.9) (35 – 75 bp) and NINA G is the reverse complement of the region preceding the zinc-binding site featured in Fig. 4.10 (approximately 200 – 250 bp according to Ammendola *et al.*, (1992)).

Using the primers NINA C and G the following gel (Fig. 4.13) shows the resulting PCR product that corresponds to a fragment of 220 bp in length.

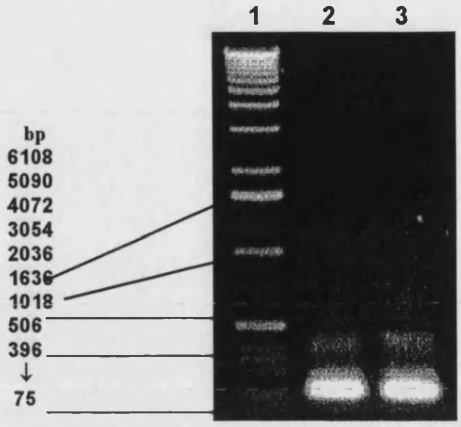


Fig. 4.13 1% agarose gel of (1)1 Kb Ladder and (2) & (3) 220 bp fragment

Therefore, as a PCR product of the required size was obtained, this was gel purified, radiolabelled and used in further hybridisation experiments as a probe.

4.3.3.2 ^{32}P Labelling and Hybridisation

Following hybridisation with the radiolabelled probe, the nitrocellulose membranes were washed using the type 2 programme. Many positively-labelled clones were identified and isolated by secondary screening.

4.3.3.3 λ Maxi Prep

One of the identified clones was amplified by infection of *E. coli* cells and the λ DNA isolated using the Maxi-prep method.

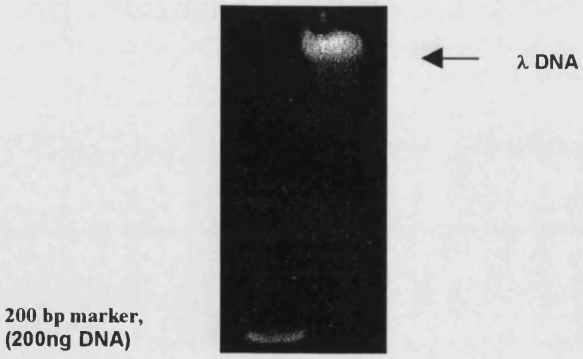


Fig. 4.14 1% agarose gel of λ DNA prepared from the positive clone.

(1)first marker of DNA Mass Ladder™, (2) λ clone ≈ 40 Kb

4.3.3.4 Sequencing

From the λ DNA, the GDH gene and its flanking regions were completely

sequenced on the sense and anti-sense strands. The sequencing strategy is shown in Fig. 4.15 and the resulting sequence in Fig. 4.16.

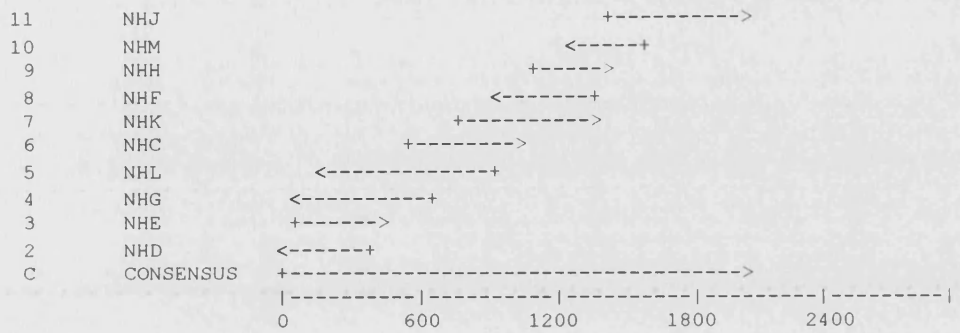


Fig. 4.15 GDH gene sequencing strategy using primers NINA C-H, J-M

4.3.3.5 Sequence analysis

In addition to acquiring the DNA sequence of the GDH gene, sequencing up and down stream of the gene revealed further information about promoter and termination signal sequences associated with it (Fig. 4.16), as well as data on the genes that flank it. Identification of the promoter and termination signal sequences was provided by sequence studies of genes isolated from the *Sulfolobus* genome by Reiter *et al.* (1988). Sequencing further along the λ clone upstream of the GDH gene revealed the end of another gene, which shared 50% sequence identity with the 3-oxoacyl-(Acyl-Carrier Protein) reductase from *Thermatoga maratima* (a thermophilic bacterium that is close to the root of the rRNA universal phylogentic tree). It also gave matches with 3-oxoacyl-reductases from *Brassica napus, Bacillus subtilis* and *Haemophilus influenza*. No ORF data could be obtained beyond the GDH gene stop site. The sequence was translated into an amino acid sequence using the translate programme of the GCG package. Observation of the amino acid sequence revealed regions of conservation noted in several sources of literature (Rossman *et al.*, 1974; Jörnvall *et al.*, 1987; Hanukoglu and Gutfinger, 1989; John *et al.*, 1994; Persson *et al.*, 1994; Edwards *et al.*, 1996), namely the putative catalytic and structural zinc-binding sites, as well as the nucleotide-binding site (Fig. 4.17).

```
 581      TAATAATTTC ATAGTATACT AGCCTGGAAT TTCTACGTAG ATTTAAAATT TTTAAGTTGT
 641  ATTGAGCATA AAATTTATAT CTGTCCCATC TTAACATAAG GTTTTAATGA AAGCTATAAT
 701   AGTGAAACCC CCAAACGCGG GAGTGCAGGT AAAGGATGTT GATGAAAAGA
 751   AATTGGACAG CTACGGGAAA ATAAAGATAA GAACAATCTA CAACGGTATT
 801   TGTGGTACTG ATAGAGAGAT AGTTAATGGT AAGTTAACAT TGTCGACTTT
 851   ACCTAAGGGA AAGGATTTCT TGGTATTGGG TCATGAGGCA ATAGGTGTTG
 901   TGGAAGAGTC TTATCATGGT TTTTCTCAAG GAGATCTGGT AATGCCAGTT
 951   AATAGGAGAG GATGCGGTAT TTGTAGAAAT TGTTTAGTTG GGAGACCAGA
1001   CTTTTGTGAA ACAGGGGAAT TCGGAGAAGC TGGAATACAC AAAATGGATG
1051   GGTTTATGAG GGAATGGTGG TACGACGACC CTAAGTATTT GGGTAAAATT
1101   CCTAAGTCAA TTGAAGATAT TGGAATTTTA GCACAACCGT TAGCAGACAT
1151   TGAGAAGTCC ATTGAGGAAA TATTGGAAGT TCAGAAAAGG GTTCCAGTGT
1201   GGACTTGTGA TGATGGAACT CTAAATTGTA GGAAAGTTCT CGTTGTTGGC
1251   ACCGGACCAA TTGGAGTCTT ATTTACTTTG CTTTTTAGAA CTTATGGACT
1301   AGAGGTTTGG ATGGCAAACA GAAGAGAACC CACTGAGGTA GAACAGACCG
1351   TTATAGAGGA AACCAAAACT AACTATTATA ATTCATCCAA TGGGTACGAC
1401   AAATTGAAAG ACTCAGTAGG TAAGTTCGAT GTAATAATAG ACGCAACGGG
1451   AGCAGATGTA AATATATTAG GTAATGTTAT TCCATTGTTA GGAAGGAACG
1501   GAGTTTTAGG TCTCTTTGGA TTTTCAACTT CTGGTTCAGT TCCATTGGAC
1551   TACAAGACAT TACAAGAAAT AGTACACACA AATAAGACGA TAATAGGTCT
1601   AGTGAATGGT CAAAAGCCCC ACTTCCAGCA AGCCGTAGTT CATTTAGCCT
1651   CTTGGAAGAC TTTATATCCT AAGGCTGCTA AGATGCTAAT TACTAAGACT
1701   GTCAGTATAA ATGATGAAAA AGAGTTACTT AAAGTGCTTA GGGAAAAGGA
1751   ACATGGAGAA ATCAAGATAA GAATATTATG GGAATAATTC AATAAAAATT
1801  TCGGTTTTTT GTGCTATCTT TTTCTGGTGT AAGGAGGTTC TAATTATGCT TGTAACAATG
1861       ATATATTTTT ACGTGAATAT GTATATAGAT TATGTATTTA
```

Fig. 4.16 Nucleotide sequence of GDH from *S. solfataricus*.

Start and Stop codons are highlighted in red. TATA box and termination sequences are highlighted in green. A putative Shine Delgarno sequence is underlined.

```
  1  MKAIIVKPPN AGVQVKDVDE KKLDSYGKIK IRTIYNGICG TDREIVNGKL
 51  TLSTLPKGKD FLVLGHEAIG VVEESYHGFS QGDLVMPVNR RGCGICRNCL
101  VGRPDFCETG EFGEAGIHKM DGFMREWWYD DPKYLGKIPK SIEDIGILAQ
151  PLADIEKSIE EILEVQKRVP VWTCDDGTLN CRKVLVVGTG PIGVLFTLLF
201  RTYGLEVWMA NRREPTEVEQ TVIEETKTNY YNSSNGYDKL KDSVGKFDVI
251  IDATGADVNI LGNVIPLLGR NGVLGLFGFS TSGSVPLDYK TLQEIVHTNK
301  TIIGLVNGQK PHFQQAVVHL ASWKTLYPKA AKMLITKTVS INDEKELLKV
351  LREKEHGEIK IRILWE*
```

Fig. 4.17 The complete amino acid sequence of GDH from *S. solfataricus*.

The key denotes the conserved regions present in the medium-chain (former long-chain) alcohol/polyol dehydrogenases (John *et al.*, 1994). Green = catalytic residues (bind catalytic zinc), Red = sequence containing conserved cysteine residues involved in binding the structural zinc, and Blue = sequence containing conserved glycine residues involved in nicotinamide binding.

A FASTA search, which looks for sequence similarity between the query sequence and any group of sequences in this case in the SWALL protein databank, revealed matches with *T. acidophilum* GDH together with a variety of alcohol dehydrogenases, two putative sorbitol dehydrogenases, 3 glutathione formaldehyde dependent and independent reductases and a formaldehyde dismutase, which with the exception of the latter are all members of the MDR family. A few examples of this search are summarized in Table 4.2.

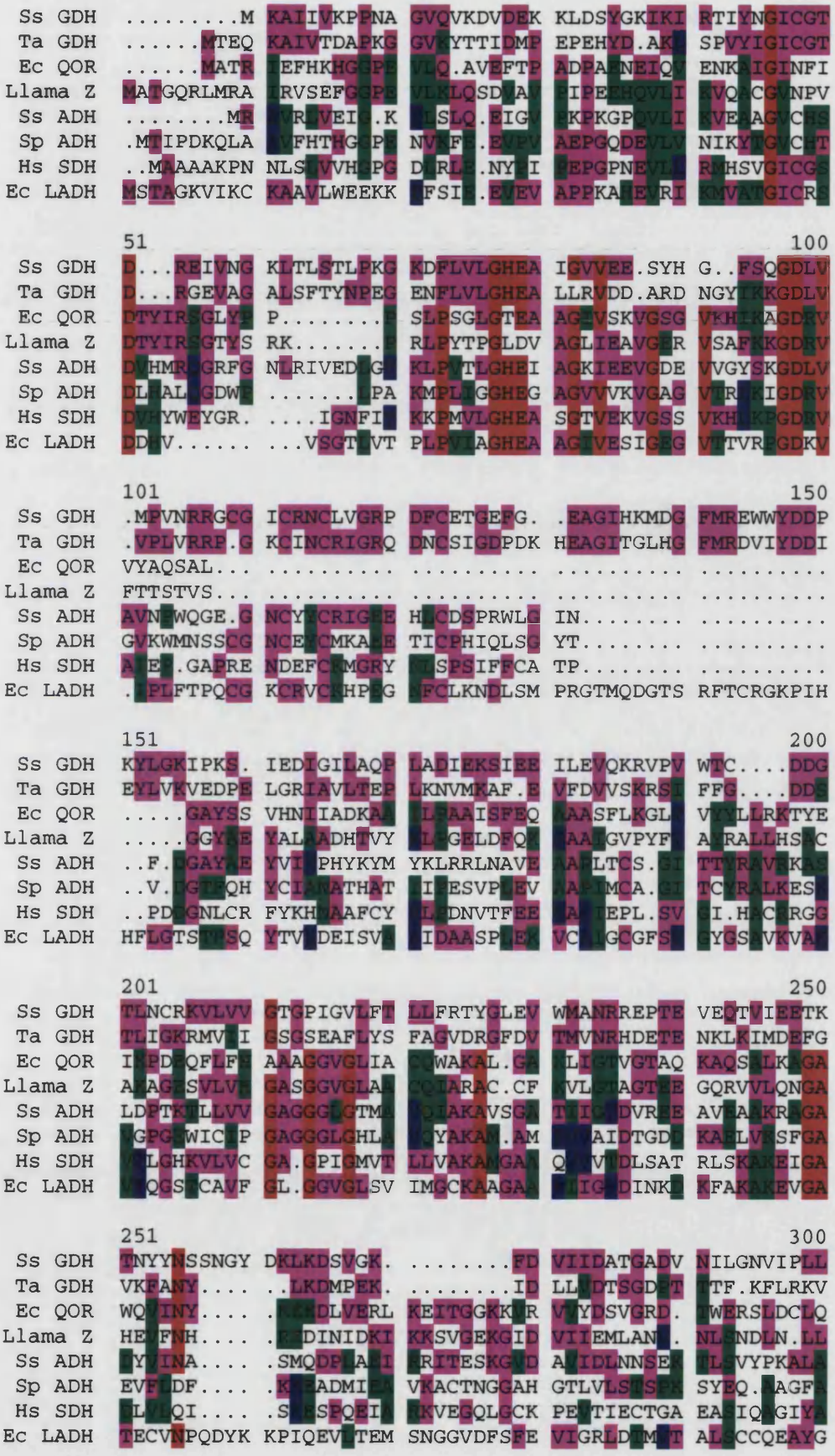
Enzyme	Source	Length (aa)	Smith - Waterman score	% Identity
Glucose 1-Dehydrogenase	*T. acidophilum*	352	651	35.3
Formaldehyde dismutase	*P. putida*	399	247	27.1
Alcohol dehydrogenase	*B sterothermophilus*	337	261	24.5
Putative sorbitol dehydrogense	*S.pombe*	360	288	24.1
NAD alcohol dehydrogenase	*B. subtilis*	378	245	22.9
Glutathione -independent reductase	*P. putida*	398	218	23.2

Table 4.2 Examples of a FASTA search through the SWALL protein databank .

Smith-Waterman score = is based on the number of matches of local regions of the sequence with all other sequences in the database (Pearson and Lipman, 1988).

A variety of sequence alignments were compiled (Fig. 4.18 - 4.20) using the PILEUP programme of the GCG package. The various colour schemes were employed in the first alignment (Fig. 4.18) are simply to show the relatedness of members of the MDR family, where *S. solfataricus* GDH was also included. This alignment was reproduced in Fig. 4.19, where only sequence identity and similarity in relation to the *S. solfataricus* GDH sequence is shown. The actual percentage sequence identities and similarities (Table 4.3) to *S. solfataricus* GDH show that the GDH from *T. acidophilum* shared the highest sequence homology, with the lowest being found with the quinone oxidoreductase from *E. coli*. The final sequence alignment (Fig. 4.20) includes *S. solfataricus* and *T. acidophilum* GDH sequences, as well as two other dual-cofactor pyridine-nucleotide-dependent GDHs from *Bacilli* (Lampel *et al.*, 1986; Heilman *et al.*, 1988). The aligment (Fig. 4.20) and the percentage sequence homology data (Table 4.4) show that, there is sequence homology between these species and implies a common evolutionary descent but with the *Bacilli* grouping more closely to eachother (Lampel *et al.*, 1986) and the archaeal GDHs also share a higher homology with each other than with the *Bacilli* sequences.

A structural alignment of the amino acid sequences from *S. solfataricus* and *T. acidophilum* was compiled (Fig. 4.21). This involved the initial submission of the sequences to a structural prediction programme, PredictProtein (Rost *et al.*, 1993; Rost *et al.*, 1994; Rost *et al.*, 1994). The amino acid sequences and their corresponding secondary structure predictions were aligned according to the fast3_t programme (Pearson and Lipman). This structural alignment was created in order to see, despite the relatively low sequence identity, whether these two archaeal GDHs share a high degree of structural homology. In view of the large number of amino acid analysis programmes available, a great deal of data can be deduced from a single amino acid sequence as already shown. Another such useful programme was employed to deduce the actual Mr of the *S. solfataricus* GDH, amino acid composition and also its isoelectric point (pI). Knowledge of the pI value and comparison of that with the pI value of *T. acidophilum* GDH would aid in the establishment of crystallisation trials, with respect to the pH of the buffers used to crystallise the *T. acidophilum* GDH (John *et al.*, 1994). In addition, knowledge and comparison of the amino acid compositions might provide information on the sequence changes necessary to allow adaptation of an enzyme to higher temperature environments.



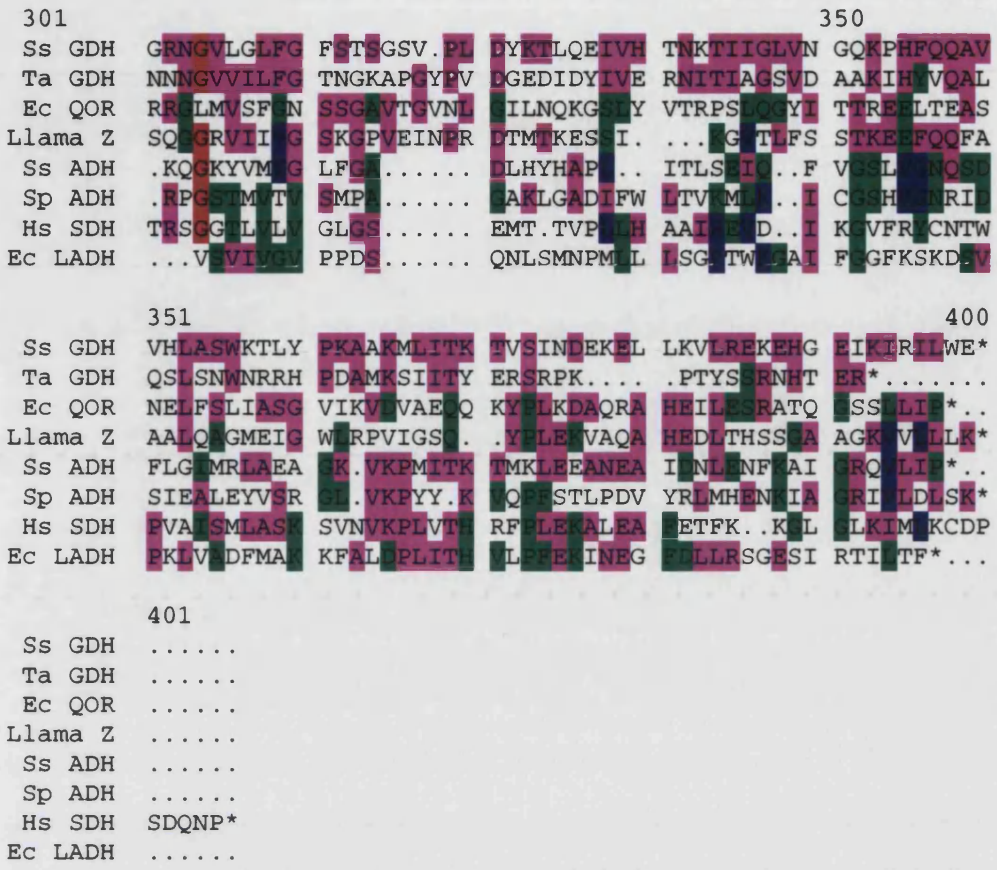
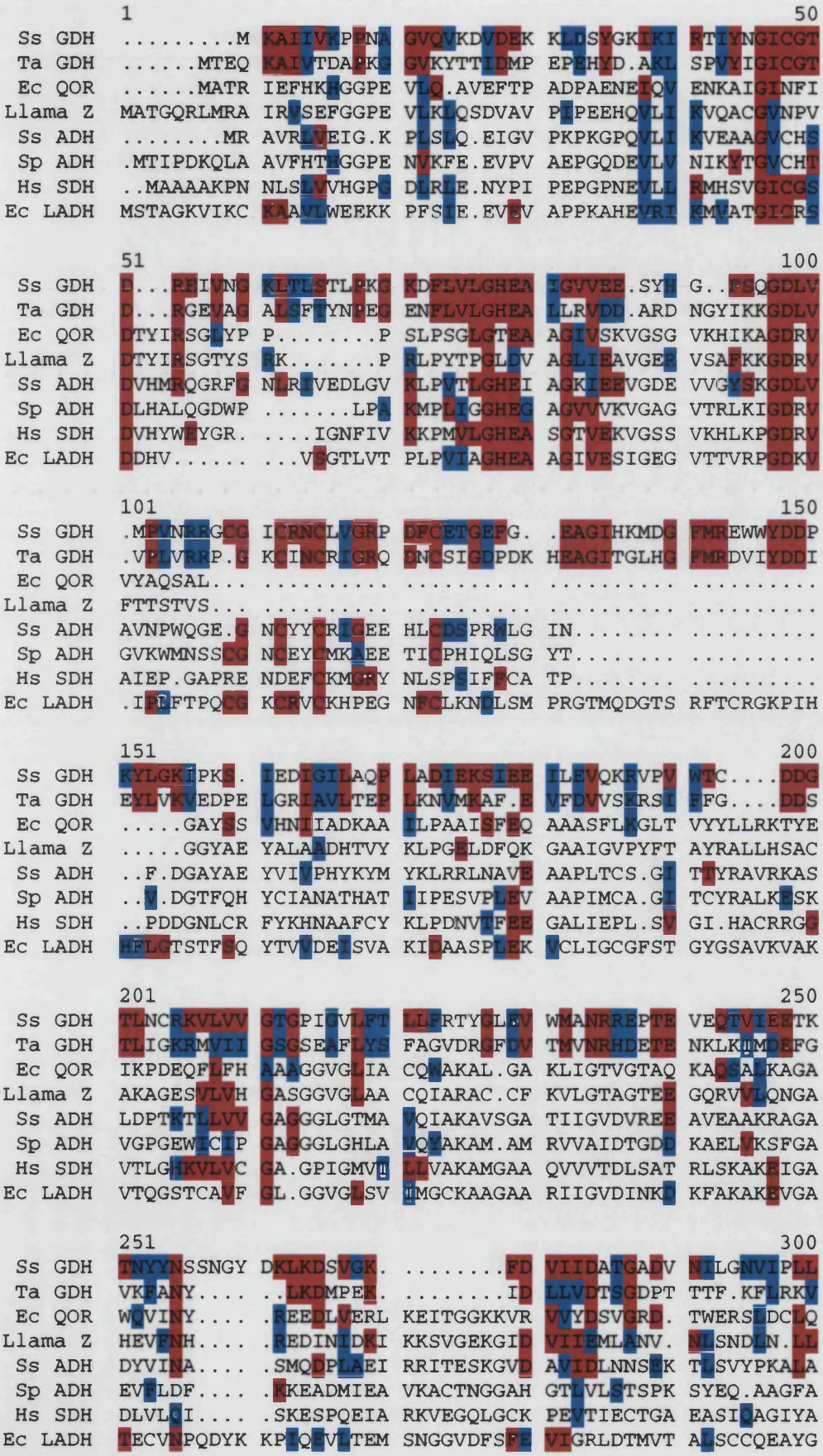


Fig. 4.18 **Depicts a sequence alignment of members of the family of MDR enzymes (constructed using pileup in GCG programme).**

***Sulfolobus* solfataricus GDH, *Thermoplasma* acidophilum GDH, *Eschericia* coli quinone oxidoreductase (QOR), Llama ζ-crystallin (Z), *S.* solfataricus alcohol dehydrogenase (ADH), *S.* pombe ADH, *Homo* sapiens Sorbitol dehydrogenase (SDH) and *Equus* caballus (Horse) liver alcohol dehydrogenase (LADH). Residues of more than 6 exhibiting sequence identity are shown in red and residues of less than 6 exhibiting sequence identity are shown in magenta, green and blue.**


1 50
Ss GDH .....M KAIIVKPNA GVQVKDVDEK KLDSYGKIKI RTIYNGICGT
Ta GDH .....MTEQ KAIVTDAPKG GVKYTTIDMP EPEHYD.AKL SPVYIGICGT
Ec QOR .....MATR IEFHKHGGPE VLQ.AVEFTP ADPAENEIQV ENKAIGINFI
Llama Z MATGQRLMRA IRVSEFGGPE VLKLQSDVAV PIPEEHQVLI KVQACGVNPV
Ss ADH .....MR AVRLYEIG.K PLSLQ.EIGV PKPKGPQVLI KVEAAGVCHS
Sp ADH .MTIPDKQLA AVFHTHGGPE NVKFE.EVPV AEPGQDEVLV NIKYTGVCHT
Hs SDH ..MAAAAKPN NLSLVVHGPS DLRLE.NYPI PEPGPNEVLL RMHSVGICGS
Ec LADH MSTAGKVIKC KAAVLWEEKK PFSIE.EVEV APPKAHEVRI KMVATGICRS
51 100
Ss GDH D...REIVNG KLTLSTLPKG KDFLVLGHEA IGVVEE.SYH G..FSQGDLV
Ta GDH D...RGEVAG ALSFTYNPEG ENFLVLGHEA LLRVDD.ARD NGYIKKGDLV
Ec QOR DTYIRSGLYP P.....P SLPSGLGTEA AGIVSKVGSG VKHIKAGDRV
Llama Z DTYIRSGTYS KK.....P RLPYTPGLDV AGLIEAVGEK VSAFKKGDRV
Ss ADH DVHMRQGRFG NLRIVEDLGV KLPVTLGHEI AGKIEEVGDE VVGYSKGDLV
Sp ADH DLHALQGDWP .....LPA KMPLIGGHEG AGVVVKVGAG VTRLKIGDRV
Hs SDH DVHYWEYGR. ....IGNFIV KKPMVLGHEA SGTVEKVGSS VKHLKPGDRV
Ec LADH DDHV..... ...VSGTLVT PLPVIAGHEA AGIVESIGEG VTTVRPGDKV
101 150
Ss GDH .MPLNRRGCG ICRNCLVGRP DFCETGEFG. .EAGIHKMDG FMREWWYDDP
Ta GDH .VPLVRRP.G KCINCRIGKQ DNCSIGDPDK HEAGITGLHG FMRDVIYDDI
Ec QOR VYAQSAL... ....
Llama Z FTTSTVS... ....
Ss ADH AVNPWQGE.G NCYYCRIGEE HLCDSPRMLG IN..... .....
Sp ADH GVKWMNSSCG NCEYCMKGEE TICPHIQLSG YT..... .....
Hs SDH AIEP.GAPRE NDEFCKMGRY NLSPSIFFCA TP..... .....
Ec LADH .IDPLFTPQCG KCRVCKHPEG NFCLKNDLSM PRGTMQDGTS RFTCRGKPIH
151 200
Ss GDH KYLGKIPKSR IEDIGILAQP LADIEKSIEE ILEVQKRVPV WTC....DDG
Ta GDH EYLVKVEDPE LGRIAVLTEP LKNVMKAF.E VFDVVSKRSI FFG....DDS
Ec QOR .....GAYSS VHNIIADKAA ILPAAISFEQ AAASFLKGLT VYYLLRKTYE
Llama Z .....GGYAE YALAADHTVY KLPGSLDFQK GAAIGVPYFT AYRALLHSAC
Ss ADH ..F.DGAYAE YVIVPHYKYM YKLRRLNAVE AAPLTCS.GI TTYRAVRKAS
Sp ADH ..V.DGTFQH YCIANATHAT IIPESVPLEV AAPIMCA.GI TCYRALKKSK
Hs SDH ..PDDGNLCR FYKHNAAFCY KLPDNVTFEE GALIEPL.SV GI.HACRRGG
Ec LADH HFLGTSTFSQ YTVVDEISVA KIDAASPLEK VCLIGCGFST GYGSAVKVAK
201 250
Ss GDH TLNCRKVLVV GTGPIGVLFT LLFRTYGLEV WMANRREPTE VEQTVIEETK
Ta GDH TLIGKRMVII GSGSEAFLYS FAGVDRIFDV TMVNRHDETE NKLKIMDFFG
Ec QOR IKPDEQFLFH AAAGGVGLIA CQWAKAL.GA KLIGTVGTAQ KAQSALKAGA
Llama Z AKAGESVLVH GASGGVGLAA CQIARAC.CF KVLGTAGTEE GQRVVLQNGA
Ss ADH LDPTKTLLVV GAGGGLGTMA VQIAKAVSGA TIIGVDVREE AVEAAKRAGA
Sp ADH VGPGEWICIP GAGGGLGHLA VQYAKAM.AM RVVAIDTGDD KAELVKSFGA
Hs SDH VTLGHKVLVC GA.GPIGMVT LLVAKAMGAA QVVVTDLSAT RLSKAKEIGA
Ec LADH VTQGSTCAVF GL.GGVGLSV IMGCKAAGAA RIIGVDINKD KFAKAKEVGA
251 300
Ss GDH TNYYNSSNGY DKLKDSVGK. .....FD VIIDATGADV NILGNVIPLL
Ta GDH VKFANY.... ..LKDMPEK. .....ID LLVDTSGDPT TTF.KFLRKV
Ec QOR WQVINY.... .REEDLVERL KEITGGKKVR VVYDSVGRDT TWERSLDCLQ
Llama Z HEVFNH.... .REDINIDKI KKSVGEKGID VIIEMLANV. NLSNDLN.LL
Ss ADH DYVINA.... .SMQDPLAEI RRITESKGVD AVIDLNNSEK TLSVYPKALA
Sp ADH EVTLDF.... .KKEADMIEA VKACTNGGAH GTLVLSTSPK SYEQ.AAGFA
Hs SDH DLVLKI.... .SKESPQEIA RKVEGQLGCK PEVTIECTGA EASIQAGIYA
Ec LADH IECVNPQDYK KPIQEVLTEM SNGGVDFSFE VIGRLDTMVT ALSCCQEAYG

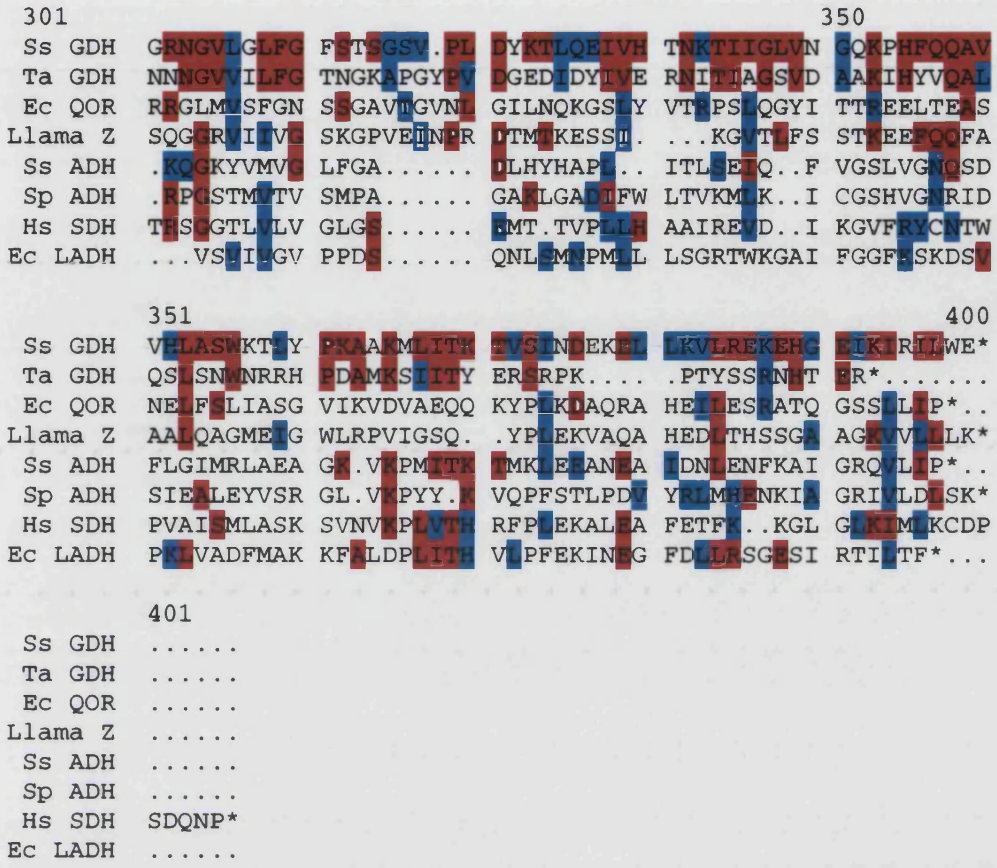


Fig. 4.19 **Sequence alignment of members of the family of MDR of enzymes with relation to *S. solfataricus* GDH.**

***S. solfataricus* GDH, *T.* acidophilum GDH, *E.* coli quinone oxidoreductase (QOR), Llama ζ-crystallin (Z), *S.* solfataricus alcohol dehydrogenase (ADH), *S.* pombe ADH, *H.*sapiens Sorbitol dehydrogenase (SDH) and *E.* caballus (Horse) liver alcohol dehydrogenase (LADH). The sequence identities and similarities are given in a tabulated form and are provided by bestfit in GCG programme. Identical residues and similar residues relative to S. s GDH are shown in red and cyan respectively.**

Enzyme	% Relation to *S. solfataricus* GDH	
	% Identity	% Similarity
T.a. GDH	35.1	57.0
E.c. QOR	16.3	42.2
Llama Z	21.5	45.8
Ss ADH	22.1	46.3
Sp ADH	20.9	43.3
Hs SDH	24.4	47.3
Ec LADH	24.7	44.1

Table 4.3 **Comparison of sequence identies and similarities of the enzymes described in Fig. 4.19.**

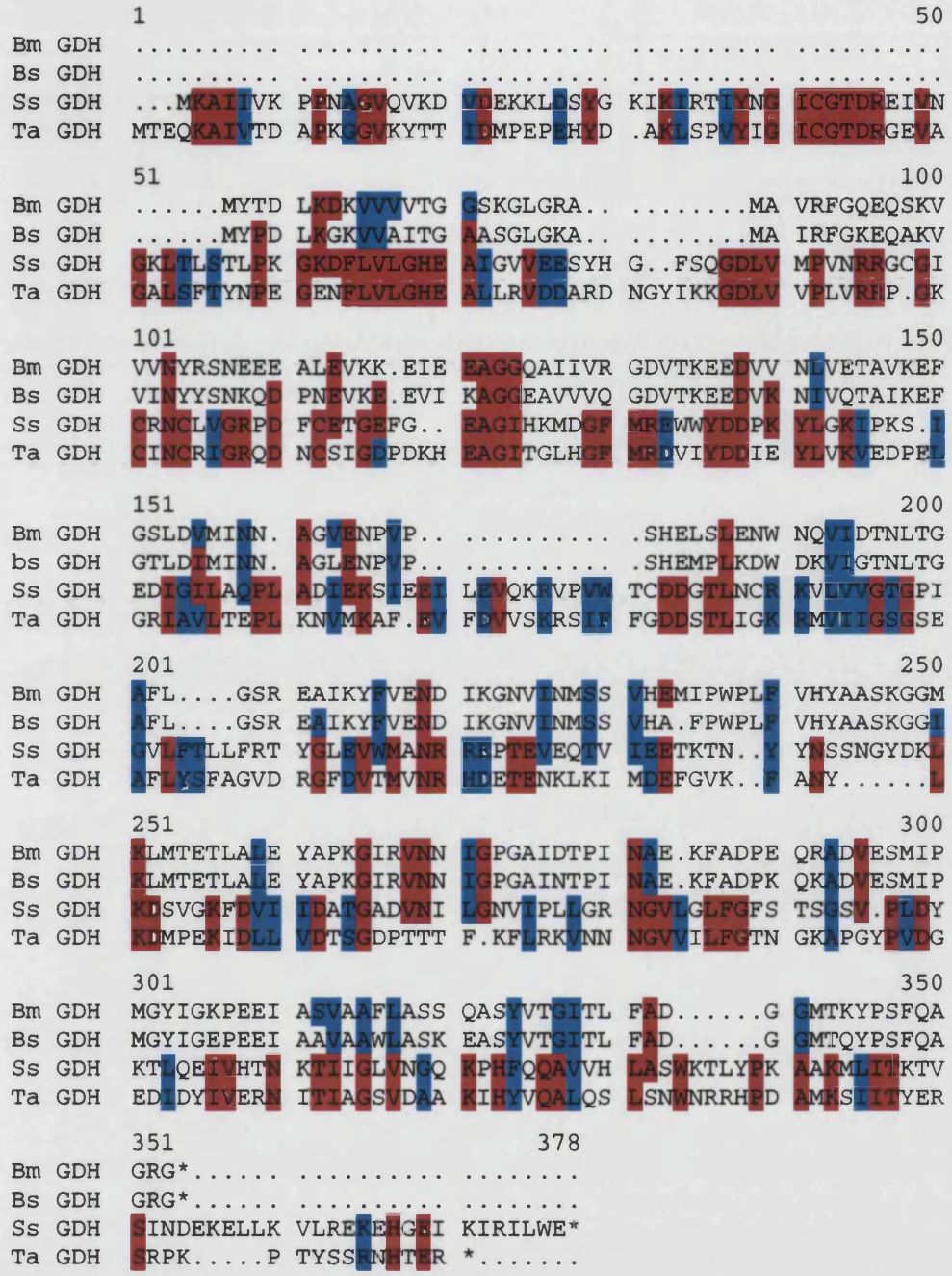


Fig. 4.20 **A sequence alignment of dual-cofactor pyrimidine-nucleotide-dependent glucose dehydrogenases.**

Bm=Bacillus megaterium **GDH, Bs=*Bacillus subtilis* GDH, Ss=*S. solfataricus* GDH and Ta=*T. acidophilum* GDH. Residues as in Fig. 4.20.**

Enzyme	% Relation to S. solfataricus GDH	
	% Identity	% Similarity
B.m. GDH	18.6	45.5
B.s. GDH	14.7	42.9
T.a. GDH	35.1	57.0

Table 4.4 **Comparison of sequence identies and similarities of the enzymes described in Fig. 4.20.**

Fig. 4.21 **Structural alignment of GDH amino acid sequences from *S. solfataricus* (black) and *T. acidophilum* (in blue) GDH.**

The numbers denote the series of amino acid residues (by a factor of ten) in the *S. solfataricus* sequence.The secondary structures are denoted by: H=helix, E=extended (sheet), L=loop and "."=no prediction is made for this residue. Sequence alignment was were compiled using the fasta3_t programme according to Pearson and Lipman, (1988). The structural predictions of the amino acid sequences were assigned according to Rost *et al.*, (1993); Rost *et al.*, (1994); Rost *et al.*, (1994).

(a)

	S. solfataricus GDH	*T. acidophilum* GDH
Molecular mass (Daltons)	**40489**	**39394**
Isoelectric point	**7.10**	**5.73**

(b)

	S. solfataricus GDH		*T. acidophilum* GDH	
Residue	**Number**	**Mole %**	**Number**	**Mole %**
Ala	13	3.5	19	5.4
Cys	7	1.9	4	1.1
Asp	21	5.7	31	8.8
Glu	27	7.4	22	6.2
Phe	11	3.0	15	4.2
Gly	36	9.8	31	8.8
His	7	1.9	8	2.3
Ile	31	8.5	26	7.4
Lys	32	8.7	24	6.8
Leu	34	9.3	23	6.5
Met	6	1.6	9	2.6
Asn	16	4.4	17	4.8
Pro	15	4.1	17	4.8
Gln	9	2.5	4	1.3
Arg	15	4.1	21	5.9
Ser	14	3.8	17	4.8
Thr	22	6.0	20	5.7
Val	33	9.0	30	8.5
Trp	6	1.6	1	0.3
Tyr	11	3.0	14	4.0

Table 4.5 a &b Chemical characteristics of GDHs from *S. solfataricus* and *T. acidophilum*.

4.2.2.6 SUB-CLONING INTO EXPRESSION VECTORS

In order to clone the GDH gene into a suitable expression vector it was first necessary to create two restriction sites at the 5' and 3' ends of the gene. Two vector/host systems were available, namely pRec7/JM109 and pET3a/BL21(DE3). These two systems differ in their mode

of expression, as the pRec7 vector system employs a synthetic *recA* promoter and the pET-3a system employs a T7-derived promoter-driven system (see Chapter 5). Therefore, it was decided to clone GDH into both these systems and continue expression work with the system that provided the most satisfactory levels of expression. Both vectors contained Nde I and BamH I sites, And therefore, after checking that no internal sites were present, PCR was used to introduce these restriction sites into the flanking regions of the GDH gene (Fig. 4.22).

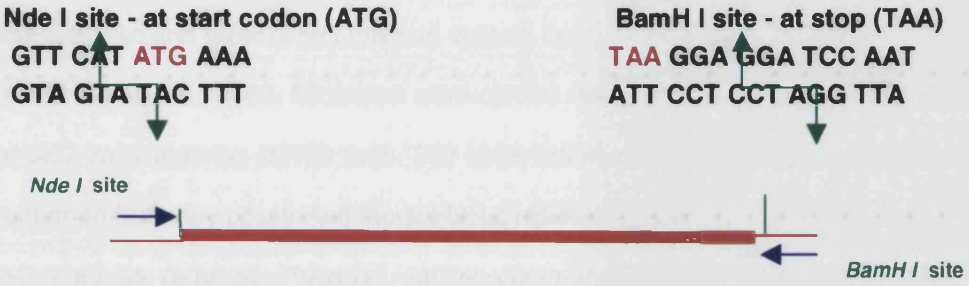


Fig. 4.22 Depiction of the inserted restriction sites required for cloning.

The PCR generated gene and both vectors were then restricted using BamH I and Nde I and following gel purification, they were ligated together. The ligation of insert (gene) and vector was confirmed by performing a second restriction digest.

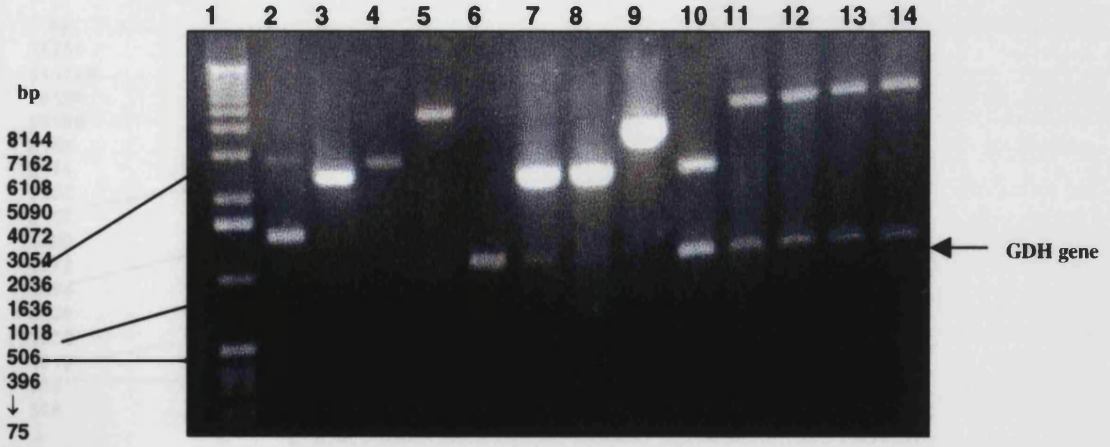


Fig. 4.23 1% Agarose Gel showing uncut recombinant vectors and doubly restricted vectors and insert, (dr).

(1) 1 Kb Ladder, (2) uncut pRec7, (3) dr pRec7, (4) uncut pET-3a, (5) dr pET-3a, (6) dr GDH, (7) - (10) dr pRec7 vectors from miniprep, (11) - (13) dr pET-3a vectors from miniprep.

The bands shown in Fig. 4.24 (lanes 2 – 5) display the characteristic migration patterns of uncut and linearised plasmid. This ability of plasmid DNA to undergo supercoiling makes it more compact than linear DNA of the same number of nucleotides and is the reason for its larger migration distance. Double restriction digests of the recombinant plasmids release the gene insert of approximately 1.1 kB. However, exceptions are seen with the pET-3a vector in lanes 7 – 9, in which the vector in lane 9 remains uncut and the vectors in lanes 7 and 8 only appear to possess a small amount of insert.

Both (pET-3a and pRec7) recombinant vectors containing insert were then transformed into the host *E.coli* strain JM109. Minipreps were carried out on four colonies from each of the pRec7 and pET-3a containing JM109 cells. The observation of growth on ampicillin containing LB-agar confirmed that they possessed the antibiotic resistance supplied by the vector and thus were transformed as required. However, before commencing expression work it was first necessary to transform the pET-3a vector plus its insert into its required expression host, namely BL21(DE3). Overexpression of gene inserts in the pET-3a vector can only occur in hosts possessing the T7 RNA polymerase gene (see chapter 5), yet to prevent plasmid instability and increase the plasmid copy number, these recombinant vectors are first cloned into T7 RNA deficient hosts. A subsequent miniprep and restriction digest of three colonies was also carried out to confirm transformation (Fig. 4.25).

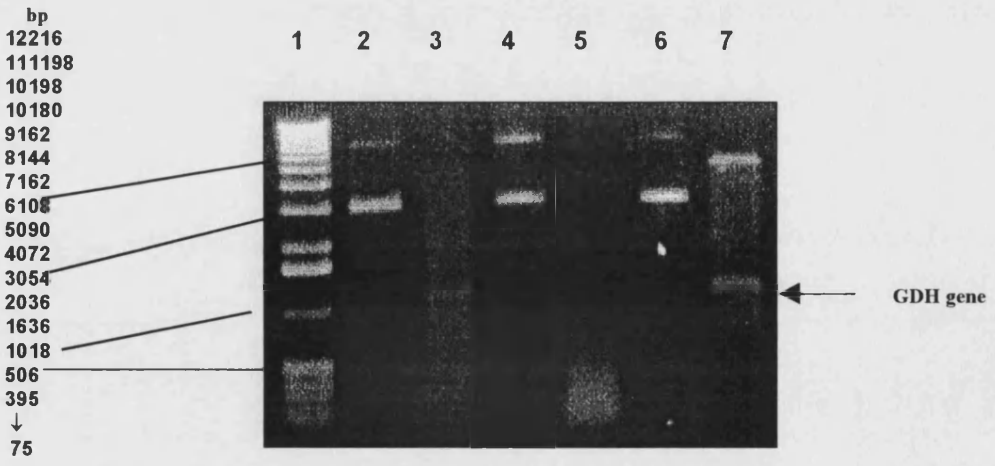


Fig. 4.25 **1% Agarose Gel to show the uncut and doubly restricted pET-3a vector and insert, (dr).**

(1)1 Kb Ladder, (2) – (7) plasmid DNA from three colonies: (2), (4), (6) uncut pET-3a and (3), (5), (7) dr pET-3a.

4.4 DISCUSSION

4.4.1 DNA SEQUENCE

The complete nucleotide sequence of the GDH gene was submitted to the EMBL Nucleotide Sequence Database (Accession number AJ012093 SSO012093). The GC content of this coding sequence is 38%, which reflects the base composition (34% G + C) of the *Sulfolobus* genome (Zillig *et* al., 1980). Start and termination signal sites are found upstream and downstream of the gene, respectively. The region upstream from the transcriptional start sites of archaeal genes contains a core promoter region that has been estimated to lie between positions –2 and –38. Alignment of archaeal sequences reveals a partially conserved A-T-rich box, which is centred approximately –26 nucleotides from the initiation site (Reiter *et al.*, 1988). Its position and A/T composition resemble the eukaryotic TATA box. Indeed, a TATA box, TTTATA, was found 27 nucleotides upstream of the start codon of the *S. solfataricus* GDH gene, as depicted in red in Fig 4.16.

The Shine Delgarno interaction is normally noted by stretches of purines located 3-9 nucleotides upstream from the start codon, and is proposed to interact with the polypyridine sequence present at the 3' terminus of each archaeal 16S rRNA (Shine and Delgarno, 1975). However, Shine Delgarno sequences for *Sulfolobus* as well as other sulfothermophiles are not very clear in that many transcripts do not possess a purine stretch while for others only two or three consecutive purines are present. In the case of the GDH gene sequence, a short stretch of purines 6 -10 nucleotides can be observed upstream of the start site. This stretch has been deemed to be a putative Shine Delgarno sequence.

Termination signals are more easily determined and are denoted by T-rich polypyrimidine sequences, as described by Reiter *et al.* (1988). They are highlighted in green in Fig. 4.16.

Sequencing further along the λ clone upstream of the GDH gene revealed the end of a putative 3-oxoacyl-(Acyl-Carrier Protein) reductase gene. No sequence ORF data could be obtained beyond the GDH gene stop site. This could be due to degradation of the λ DNA and would require another DNA preparation

4.4.2 AMINO ACID SEQUENCE

The complete sequencing of the GDH gene allowed comparisons to be made with other medium-chain (former long-chain) dehydrogenases. Fig. 4.18 shows the relationship of representatives of this family with each other. Although there are only a few regions of absolute identity across all the representative members, which includes the putative structural zinc-binding and nucleotide-binding sites, there are also many instances of identities within the group (i.e. less than six members showing sequence identity. Fig. 4.19 and Table 4.3 show the relationship between the deduced amino acid sequence of the *S. solfataricus* GDH and other members of this family.

The alignment of dual-cofactor pyridine-nucleotide-dependent glucose dehydrogenases (Fig. 4.20) shows that, although the GDHs of Archaea and Bacilli both catalyse the oxidation of glucose and the concomitant reduction of $NAD(P)^+$, they share very low sequence identities (Table 4.4). However, this is reflected in their groupings with other dehydrogenases within the pyridine-nucleotide-dependent alcohol/polyol/sugar dehydrogenase/reductase superfamily. The archaeal GDHs group with the medium-chain (former long-chain) frequently zinc-containing dehydrogenases, and the GDHs from the Bacilli lie within the short-chain, non-zinc-containing branch of this superfamily (Jörnvall *et al.*, 1984), which also groups with zinc-free ribitol dehydrogenases and the *Drosophila* alcohol dehydrogenase.

The secondary structure prediction alignment of *S.* solfataricus GDH and *T. acidophilum* (Fig. 4.21) reveals areas of identical secondary structure with the exception of the C-terminal portion, where *T. acidophilum* GDH possesses a rather extended helical region followed by a loop structure, whereas the *S. solfataricus* GDH consists of an extended sheet region followed by a short helical section and ends in a mixture of loop/extended sheet which extends further than the *T. acidophilum* GDH amino acid sequence. However, to determine more precisely what differences occur in the structures of the GDHs it is necessary to create a protein model of *S. solfataricus* GDH based on another known structure or to obtain its crystal structure

The amino acid sequence alignment of *S. solfataricus* GDH, *T. acidophilum* GDH, *E. coli* quinone oxidoreductase (QOR), Llama ζ-crystallin (Z), *S. solfataricus* alcohol dehydrogenase (ADH), *S. pombe* ADH, *H.* sapiens Sorbitol dehydrogenase (SDH) and *E.*

caballus (horse) liver alcohol dehydrogenase (LADH) (Fig 4.19) revealed respective identities of 35%, 16%, 22%, 22%, 21%, 24% and 24.7% (Table 4.3). The relatively low sequence identity with the *T. acidophilum* GDH (which did actually correspond to a 56.1% DNA sequence identity of the internal region of the GDH genes- result not shown) was not sufficient for successful hybridisation experiments. However, on closer examination of the sequences it can be seen that in fact many specific segments have been conserved, suggesting that *S. solfataricus* GDH belongs to the MDR family (Jörnvall *et al.*, 1987).

4.4.2.1 Structural Zinc-binding site

This site is denoted specifically in the sequence by the presence of four cysteine residues, 93 - 107 aa (only three in *T. acidophilum* GDH). The cysteines are fundamental in the co-ordinate binding of a zinc atom. Although, it still remains to be tested whether the *S. solfataricus* GDH does in fact contain zinc atoms.

4.4.2.2 Catalytic Zinc-binding site

The ligands to the catalytic zinc, two cysteine residues (Cys46, Cys174) and one histidine residue (His67), are highly conserved throughout the alcohol dehydrogenase family (Jörnvall *et al.*, 1987). In *T. acidophilum* GDH, the ligands involved are Cys46, Glu155 and His67, Glu155 occuring in a structurally-equivalent position to the Cys174 of LADH Therefore, in the apo-enzyme, the tetra-co-ordinate system is completed with a water molecule (John *et al.*, 1994). Sorbitol dehydrogenase also appears to possess a glutamate (Glu174) instead of a second Cys. In *S. solfataricus* GDH these residues are found at Cys39, Cys174 and His66 (Fig. 4.17).

4.4.2.3 Nucleotide-binding site

The *S. solfataricus* GDH, like that of *T. acidophilum* as well as many other archaeal dehyrogenases (Danson, 1993), is dual-cofactor specific. It preferentially utilises $NADP^+$, binding NAD^+ with lower affinity (data shown in chapter 6). Previous studies concerning the nucleotide-binding domain have associated the GxGxxG sequence motif in conjunction with a negatively-charged residue at the end of the βB strand, with NAD^+ specificity (Wierenga *et al.*, 1985; Wierenga *et al.*, 1986; Hanukoglu and Gutfinger, 1989; Scrutton *et al.*, 1990; Baker *et al.*, 1992a), On the other hand, the GxGxxA sequence motif plus a positively-charged residue at the

end of the βB strand, confer $NADP^+$ specificity.

On viewing the amino acid sequence in Fig 4.17, it can be seen that *S. solfataricus* GDH possesses the GxGxxG sequence motif associated with NAD^+ specificity. However, it also possesses an asparagine residue at the end of βB. Asn211 of *S. solfataricus* GDH is structurally equivalent to Asn215 of *T. acidophilum* GDH, Asp223 of LADH and Arg218 of glutathione reductase. The presence of Asn211 would enable a weak interaction to be made with the adenosine 2-hydroxyl. The binding of $NADP^+$ requires stabilization of the adenosine 2-phosphate. In *S. solfataricus* GDH this may arise from Arg213. However, in order to be able to propose a mechanism of nucleotide binding it is first necessary to model the *S. solfataricus* GDH amino acid sequence on the *T. acidophilum* GDH structure (see chapter 8) or indeed obtain a crystal structure of *S. solfataricus* GDH. This information will tell us which residues lie in the active site and how they may interact with the cofactor.

4.4.2.4 Further analysis

Further information about an enzyme's properties allow for further comparison between with the *S. solfataricus* GDH and the *T. acidophilum* GDH. For instance submission of these sequences into PEPTIDESORT in the GCG programme (Table 4.5a) showed that although these enzymes are similar in size they differ greatly in isoelectric points and this would also explain the slight difference in purification strategies in which the *T. acidophilum* GDH employed an anion exchange step at pH 7.0, whereas the *S. solfataricus* GDH was successfully bound to cation exchange and anion exchange chromatography at pH 5.5 and 8.5 respectively. This may also explain their difference in their requirements for binding to Matrex Gel Red A in which *S. solfataricus* GDH employs a lower pH buffer. Also determined by this programme are the total number of amino acids per subunit and their respective mole percents (Table 4.5b). Comparison of these data between the archaeal GDHs (*S. solfataricus and T. acidophilum*) reveals an increase in the number of aliphatic residues with the exception of alanine. These amino acid residues have no reactive groups on their side chains, only inert methylene ($-CH_2$) and methyl ($-CH_3$) groups. Their ability to interact much more favourably with each other and other nonpolar atoms than with water leads to the creation nonpolar environments. This phenomenon is known as the hydrophobic effect and is one of the main factors in stabilising the folded conformations of proteins. These data therefore suggest one of the ways in which the GDH from *S. solfataricus*

has managed to adapt to function in such a high temperature environment in comparison to the *T. acidophilum* GDH. An increase in the hydrophobic residues could lead to an increase in compactness of the enzyme due to the reduction in internal cavities, increased internal packing through the formation of isoleucine clusters and greater complementarity at the subunit interface due to more hydrophobic contacts. The appearance of these features are thought to bring about greater rigidity in the protein, thus increasing its thermostabilty. Also visible is an increase in Tryptophan residues from one in *T. acidophilum* GDH to six in *S. solfataricus* GDH, which is unusual and therefore significant as this residue does not occur very frequently, so proteins often have only one or a few Trp residues. The aromatic, nonpolar nature of this residue may also contribute to increased internal packing, although observation of the overall number of aromatic residues in both GDHs reveals that, they are virtually unchanged.

The presence of thermolabile amino acid residues is thought to limit enzyme thermostabilty due to their tendency to undergo oxidation (Cys, Met) or deamidation (Asn, Gln) at high temperatures. There is a slight reduction in Met residues, the number of Asn residues is virtually unchanged and there are actually increased numbers of Cys and Gln residues in the *S. solfataricus* GDH. However, knowledge of the position of these residues in the tertiary structure of *S. solfataricus* GDH would increase the significance of these data. For instance, if the majority were found to be buried in the solvent-free interior of the protein, this would render them less susceptible to covalent modification. Also the higher incidence of Cys residues in *S. solfataricus* GDH can be explained by their use in the co-ordinating of zinc atoms. An increase in other amino acid interactions thought to participate to protein thermostabilty (Perutz and Raidt, 1975), cannot be deduced simply from knowledge of the sequence and composition of amino acids of a protein and instead requires in depth knowledge of the protein structure.

5 Chapter Five

Recombinant expression & purification of Glucose dehydrogenase

5.1 INTRODUCTION

In order to obtain a high level of expression of thermophilic GDH, it was thought desirable to express the recombinant protein in *E. coli*. There are three advantages of using an *E. coli* system as opposed to expression in the native organism. Firstly, the rate of growth of *E. coli* is significantly faster than that of *S. solfataricus* even when the latter is grown under optimal conditions. A culture of *S. solfataricus* would require a week to ten days to reach the equivalent cell density attained by an overnight culture of *E. coli*. The second advantage of employing the *E. coli* system is the ease of culturing, as not only is *S. solfataricus* difficult to grow, it also does not possess a suitable host/vector system for overexpression of proteins and thirdly the *E. coli* system has the advantage of facilitating the purification of the desired protein. The differences in the properties of an expressed thermophilic protein from those of *E. coli* can be exploited such that the addition of a high temperature step would cause the denaturation of the mesophilic *E. coli* proteins, leaving the expressed thermophilic enzyme in solution.

High levels of expression were needed to produce sufficient purified enzyme for characterization and crystallization. Many recent techniques such as placing genes of interest downstream of strong promoters for efficient transcription or their insertion into multi-copy plasmids in order to elevate gene dosage have brought about the over production of proteins.

Some vectors with portable promoters have been constructed, examples including the *lac* operon (Robert *et al.*, 1979), *trp* operon (Hallewell and Emtage, 1980), λ phage (Remaut *et al.*, 1981; Tsurimoto *et al.*, 1982; Moffat and Studier, 1986: Rosenburg *et al*,. 1987: Studier *et al*,. 1990) and *recA* (Feinstein *et al.*, 1983; Shirakawa *et al.*, 1984; Kurz *et al.*, 1997). For a promoter to

function ideally it should be able to over-express the inserted gene, function only under inducing conditions and have a simple method of induction so that large-scale operation is possible.

Vector systems containing the promoters of the *E. coli recA* gene and the bacteriophage T7 have been chosen because they fulfill the above requirements and were also currently employed and available for use in our research group.

pRec7/ Nde I

The vector pRec7 is based on the vector pARC306A (Kurz *et al.*, 1997) with the simple addition of an Nde I site. The pUC-based plasmid, pARC306A, has a synthetic *recA* promoter, a synthetic enhancer region based on the sequence of the T7 promoter 10 leader (a Shine-Delgarno sequence), a restriction site cluster, and transcription terminators at both ends of the expression cassette. The presence of the Nde I site (5'- CATATG -3') allows for convenient cloning of the start codon (ATG) of the gene at an optimal distance from the Shine-Delgarno sequence (Gerike *et al.*, 1997).

The gene product of *recA* performs many essential roles in genetic recombination and in the SOS regulatory response in *E. coli.* (Gottesman, 1981; Radding, 1981). Under normal growth conditions the *recA* gene together with a set of unlinked genes are repressed by the *lexA*-gene product. However, this does however allow a limited amount of RecA protein production (Shirakawa *et al.*, 1984; Little and Mount, 1982). The SOS regulatory response is displayed in response to conditions that damage DNA or inhibit DNA replication. Initiation of the SOS response can be brought about by DNA damaging agents such as ultraviolet irradiation, alkylating agents and crosslinking agents such as mitomycin C (Iyer *et al.*, 1964). It can also be brought about by agents that inhibit replication such as nalidixic acid (Feinstein *et al.*, 1983) and thymine deprivation (Little and Mount, 1982). The SOS response itself involves processes including enhanced capacity for DNA repair and mutagenesis (Humayun, 1998), inhibition of cell division and prophage induction.

An induction signal is believed to be the presence of exposed single stranded DNA at

replication forks brought about by aborted replication. These sites are then susceptible to binding by the RecA protein. The DNA-bound RecA protein then undergoes a conformational transition to RecA*, the SOS-active form of the RecA protein. The resulting RecA*-DNA complex has co-protease activity towards the two SOS proteins; LexA, which inactivates its repressor activity and induces the SOS target genes allowing their elevated expression; and UmuD, which activates it to the UmuD' form required for its role in SOS replication (Little and Mount, 1982; Humayun, 1998)

Therefore addition of an SOS-inducing agent such as nalidixic acid stimulates basal levels of RecA protein to cleave its repressor and thus induce itself. This results in a majority of RecA proteins in the cell (Gudas and Pardee, 1976). This indicates that a combination of the *recA* promoter and a ribosome binding site provides an efficient system for inducible expression (Feinstein *et al.*, 1983; Shirakawa *et al.*, 1984; Kurz *et al.*, 1997). However, it is required that the host strain be *recA* positive in order for this system to operate (Shirakawa *et al.*, 1984).

pET-3a

The pET-3a vector is based on the plasmid pBR322 and employs a T7-derived promoter-driven system (Moffat and Studier, 1986; Rosenburg *et al.*, 1987; Studier *et al.*, 1990). The gene of interest is cloned into the pET-3a plasmid downstream of a T7 promoter. Expression is achieved by the presence of a source of T7 RNA polymerase in the host cell. However, to prevent plasmid instability due to the production of potentially host-toxic proteins, the plasmids are initially transformed into a host that does not possess the T7 RNA polymerase gene and so can be termed as effectively "switched off". None-the-less, some degree of transcription occurs in the uninduced state. On establishing themselves in the T7 RNA polymerase deficient host, they are then transferred into the expression host. This host possesses a copy of the T7 RNA polymerase gene that it is under *lacUV5* control (λDE3 lysogen) within its chromosomal DNA. Addition of IPTG induces expression by acting on the *lacI* gene product (the lac repressor).

This chapter outlines the growth and induction of cells expressing the recombinant GDH gene from *S. solfataricus* and describes the purification of the recombinant protein based on a heat step and methods employed in purification of the native protein (Chapter 3).

5.2 METHODS AND MATERIALS

5.2.1 Reagents

Nalidixic acid, lactate dehydrogenase and Reactive Red (an alternative affinity gel medium to Matrex Red gel A but which works on the same principle) were all purchased from Sigma-Aldrich Company Ltd., Poole, Dorset, U.K. and IPTG was obtained from Calbiochem, La Jolla, CA, USA. For details of other reagents and equipment please refer to section 3.2. Details of vectors are supplied in section 4.2.

5.2.2 Cell culture and induction

A colony from a plate culture (agar supplemented with 100 μg ampicillin/ml) prepared as described in section 4.2 was used to inoculate 5 ml of sterile LB broth containing 100 μg ampicillin/ml. The culture was then placed in a shaking incubator at 37°C for approximately 8 h. 2 ml of this culture was then used to inoculate 100 ml of sterile LB broth containing ampicillin. This was then allowed to incubate in a shaking incubator at 37°C overnight. The following morning the entire 100 ml culture was used to inoculate 1 L of sterile LB broth plus 100 μg ampicillin/ml. This was incubated in a shaking incubator at 37°C until it reached an OD_{600} of 0.6. At this point, the culture was then induced with 50 μg/ml (final concentration) of nalidixic acid if it was a JM109 culture possessing the pRec7/Nde I vector, or with 0.4 mM (final concentration) IPTG if it was a BL21(DE3) culture transformed with pET-3a vector. The induced culture was maintained at the incubation conditions until the following morning, when it was centrifuged at 18,000 x g for 30 min to produce a cell pellet that could be stored at -20°C until required.

5.2.3 Cell lysis

Approximately 10 g of frozen or fresh cell paste was resuspended in 15 ml of 50 mM MES pH 5.5, 20 mM $MgCl_2$, 1 mM PMSF, 1% Triton X-100 and 10 mg lysozyme/ml, and the suspension was incubated in a water bath at 37°C for 60 min with occasional stirring. The resulting viscous mixture was then sonicated on ice with five 30 s bursts with 30 s pauses or until the mixture appeared to have gained fluidity and was slightly darker in colour. Following this, the mixture was

centrifuged at 18,000 x g for 30 min. The cell debris was discarded and the supernatant was retained for purification of the GDH enzyme.

5.2.4 Heat treatment step

The supernatant described in section 5.2.4 was placed in a 50 ml Falcon tube and was incubated at 80°C in a water bath for 30 min. The resulting mixture of denatured mesophilic proteins and soluble fraction were then transferred to Sorvall tubes and centrifuged at 18,000 x g for 30 min. The supernatant containing the soluble *S. solfataricus* GDH was retained for further purification.

5.2.5 Affinity chromatography

The supernatant described in section 5.2.4 was applied to an affinity red gel column, which was prepared and subjected to a salt gradient identical to that described in section 3.2.2.1.

5.2.6 Preparation of whole cells for SDS-PAGE

A 1 ml aliquot of cell culture was centrifuged at 13,000 x g for 30 s in a bench top microcentrifuge. The resulting cell pellet was resuspended in 50 µl of 50 mM MES pH 5.5, 20 mM $MgCl_2$ buffer and 50 µl of SDS loading buffer and heated at 100°C for 5 min. This mixture was then loaded on to an SDS-PAGE,

5.2.7 Native molecular weight determination

The sample of GDH purified in section 5.25 was applied to a gel filtration column (see section 2.3.4) as a polishing and buffer exchange step in 50 mM Tris/HCl pH 7, 20 mM $MgCl_2$. Two consecutive chromatography runs were then carried out on the gel filtration column in the same buffer. The first employed 2 ml of 1.5 mg/ml of pure lactate dehydrogenase as a size marker and the second employed 2 ml of the purified and polished GDH at 2 mg/ml.

5.3 RESULTS

5.3.1 Expression

BL21 and JM109 transformed cells were both grown to an OD_{600} reading of 0.6 and then induced with IPTG and nalidixic acid, respectively. Samples were collected at various time points after induction in order to deduce the time required to produce the highest levels of expression of GDH (Fig. 5.1). This showed that the desired protein (GDH) was being induced to dectectable levels after just 2 h in the JM109/pRec7 system and the system employing the pET-3a shows evidence of gene expression without undergoing induction. In addition, induction does not appear to increase the level of GDH protein produced. Therefore, there was a suggestion of a higher level of expression in the pRec7 system.

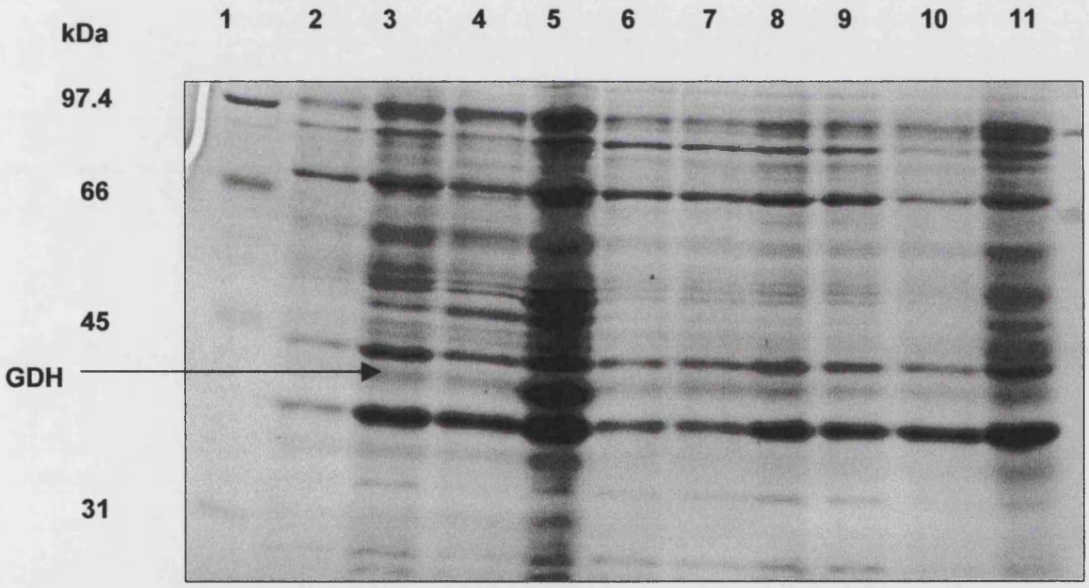


Fig. 5.1 SDS-PAGE of whole cell extracts of the JM109/pRec7 system and the BL21/pET-3a system after various induction periods.

(1) Low range marker. pRec7: (2) 0 hours, (3) 2 hours, (4) 4 hours, (5) overnight. pET-3a :- (6) 0 hours, (7) 1 hour, (8) 2 hours, (9) 4 hours, (10) 5 hours, (11) overnight.

Different temperatures for induction were also employed to observe any improvement on expression levels (Fig. 5.2). GDH induction with the JM109/pRec7 system after 24 h at 37°C produced the highest levels and induction using this system at lower temperatures failed to produce such substantial amounts of GDH protein. GDH induction in the BL21/pET-3a system was detectable, yet inferior to that of the JM109/pRec7 system. However, the level of induction was improved by incubation at 25°C.

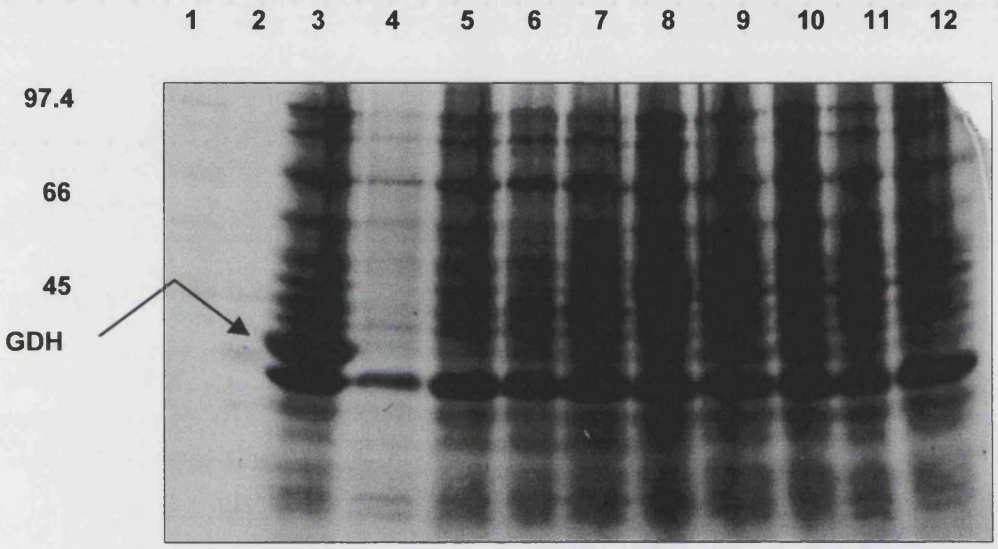


Fig. 5.2 SDS-PAGE of whole cell extracts of JM109/pRec7 system and the BL21/pET-3a system showing periods after induction at different temperatures.

(1) Low range SDS marker, (2) Native GDH sample from approach 1 (Fig. 3.3), (3) pRec7 in JM109 24 h at 37°C, (4) untransformed JM109 cells 24 h at 37°C, (5) pRec7 in JM109 5 h at 37°C, (6) pRec7 in JM109 24 h at 25°C, (7) pRec7 in JM109 24 h at 30°C, (8) untransformed BL21 cells 24 h at 37°C, (9) pET-3a in BL21 24 h at 20°C, (10) pET-3a in BL21 24 h at 25°C, (11) pET-3a in BL21 24 h at 30°C, (12) pET-3a in BL21 24 h at 37°C

5.3.2 Recombinant Purification

On deciding to employ the JM109/pRec7 system for GDH induction, it was then necessary

to purify the expressed protein in order to carry out the characterisation studies. The recombinant GDH purification involved a heat step to remove all the mesophilic host *E. coli* proteins and an affinity chromatography step. The efficiency of these steps is displayed in the following purification table (Table 5.1) and an SDS-PAGE (Fig. 5.3).

STEP	TOTAL ACTIVITY (μmoles/min)	TOTAL PROTEIN (mg)	VOLUME (ml)	RECOVERY (%)	SPECIFIC ACTIVITY (μmoles/min/mg)	FOLD PURIFICATION
Cell Free Extract	1083	531	18	-	2.04	-
Heat Step	1535	67	15	142	22.8	11.2
Matrex Gel Red A	856	8.3	21	79	103	50.4

Table 5.1 **Purification of recombinant *S. solfataricus* GDH expressed in *E. coli* strain JM109 using the pRec7 vector system.**

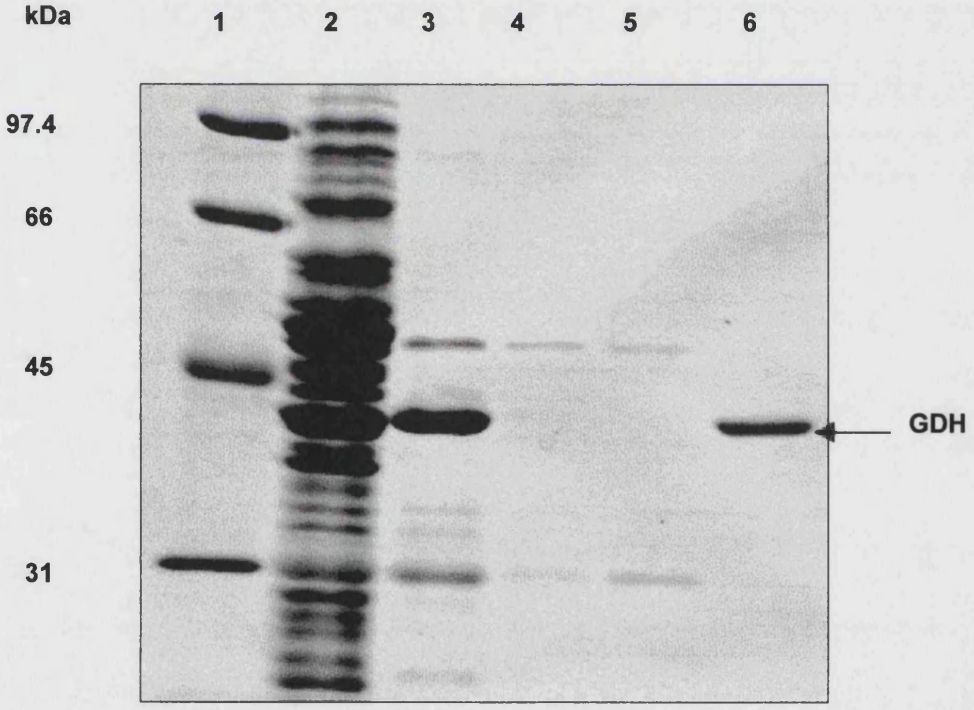


Fig 5.3 **SDS-PAGE analysis of the purification of recombinant *S. solfataricus* GDH expressed in *E. coli* strain JM109 using the pRec7 vector system.**

(1) Low Range SDS Markers. (2) Cell Extract, (3) Heat Treatment. Matrex Gel Red A Chromatography Step :-(4) Pooled flow through, (5) Pooled wash, (6) Pooled eluted GDH.

5.3.3 Native molecular weight determination

LDH is a tetrameric enzyme and has a native molecular weight of 140 kDa. On subjecting it to gel filtration chromatography (Superdex 200), under non-denaturing conditions, it gave an elution volume of 63 ml. A second gel filtration run using purified GDH on the same column and with the same buffer, gave an elution volume of 48 ml (chromatography profiles not shown). The smaller elution volume achieved with the GDH protein would therefore suggest that it has a native molecular weight greater than 140 kDa. Therefore, taking into account the subunit molecular weight of GDH, which was determined to be 41 kDa, the GDH protein must be a tetramer of approximately 165 kDa in order to elute before the 140 kDa LDH.

5.4 DISCUSSION

5.4.1 Expression

The GDH gene was cloned into two vector systems – pRec/JM109 and pET-3a/BL21(DE3). Both systems yielded the expression of a protein of approximate Mr 40 kDa. The fact that it is also seen in the pET-3a/BL21(DE3) system confirms that this is the GDH protein and not *rec*A gene product, which also has an Mr of 40 kDa (Gudas *et* al., 1977). This was further confirmed by co-electrophoresing a sample of purified native GDH in a neighbouring lane on SDS-PAGE (Fig. 5.2, lane 2). The failure to purify the native sample to homogeneity and its storage at 4°C for several months had caused some degradation. However, a protein band in the native sample had migrated to exactly the same distance as the expressed protein. In view of the nature of the pRec7/Nde I expression system, one would normally expect to see co-expression of the *rec*A gene product. However, the *E. coli* strain used is actually a *rec*A minus strain and according to the literature should not allow induction (Shirakawa *et al.*, 1984). Although it has been shown that expression of a psychrophilic citrate synthase was also achievable with the *rec*A minus *E. coli* strain MOB154 (Gerike *et al.*, 1997), it must also be noted that the level of expression was even higher in a *rec*A positive *E. coli* strain (W620), due to the presence of an active RecA protein.

It was decided that, based on the appearance of slightly higher expression yields with the pRec/JM109 system, further work concerning protein expression would continue with this system.

In addition, it was actually easier to work with in that it did not require a second transformation step. Work on varying the temperature and duration of induction also revealed that pRec7/GDH/JM109 cells required an overnight incubation at 37°C in order to achieve the best yields of GDH.

5.4.2 Recombinant purification

The recombinant GDH protein was successfully purified using fewer steps than with the native enzyme. This was achieved by the higher level of expression of GDH in the recombinant *E. coli* host, which was increased by a factor of 6.2 and may be increased even further by expression in a *rec*A positive host (as explained previously). Also the ability to use a heat step in the recombinant purification provided a very thorough purification step and additionally showed a 42% increase in total activity of the sample. Therefore, treatment of the cell extract with heat could possibly cause some or all the GDH that had not folded properly to refold and attain an active conformation, as seen with the glutamate dehydrogenase isolated from *P. furiosus* by Robb, Park, & Adams (1992). This is perhaps achieved by simply placing them in their "native", hot, environment. As it must be noted that translated thermophilic proteins may not be suited to folding properly at mesophilic temperatures. The affinity chromatography step also constitutes a substantial and effective purification step, creating a 124-fold and 50-fold purification in the native and recombinant protein sample respectively.

Overall the recombinant purification delivers the higher efficiency approach as expected. The higher specific activity of GDH in the cell extract due to increased expression and the ability of using a heat step due to the thermolabilty of the host proteins means that GDH can be purified to homogeneity (specific activity of 103 μmoles/min/mg) over two purification steps and a 50-fold purification. This is as opposed to three steps in the native purification, which require a 124-fold purification due to the lower expression levels and still only achieves a specific activity of 41 μmoles/min/mg. The difference in specific activities achieved with these purifications does not coincide with their apparent purity on SDS-PAGE. However, silver-staining of the SDS-PAGE would have more accurately revealed the existence of any other contaminating protein bands. Another possibility is the inactivation of the native GDH, during the period of the final purification step, which employed unfavourable conditions, namely the low pH (pH 5.5) and the high salt (1.5 M

NaCl).

5.4.3 Native molecular weight of GDH

The tetrameric composition of the *S. solfataricus* GDH is in accordance with many other pyridine-nucleotide-dependent dehydrogenases (see section 1.5.1) including the GDH from the archaeal thermophile *T. acidophilum* (Smith *et al.*, 1989). However, it does differ from the dimeric GDH isolated from the fellow archaeal hyperthermophile *T. tenax* (Siebers *et al.*, 1997), which is actually closer in phylogenetic and temperature range location to *Sulfolobus* than is the genus *Thermoplasma* (see Fig.1.2).

6 Chapter Six

Characterisation of the recombinant Glucose dehydrogenase

6.1 INTRODUCTION

Following the work carried out to clone and sequence the GDH gene and express the protein in *E. coli* (chapters 3, 4 and 5), it was then necessary to characterise the recombinant enzyme. This is initially very important in order to deduce whether, despite undergoing recombinant expression, it still possesses exactly the same or almost identical properties to the native enzyme. Secondly, the application of this enzyme in an *in vitro* system (see chapter 7) requires knowledge of its functional parameters in order to view the suitability of the enzyme and/or to make necessary changes to the system to facilitate its optimum function. Thirdly and finally, fully characterising this thermostable enzyme will provide more general information on protein stability at elevated temperatures and indeed on the stability and catalytic ability of thermophilic aldose dehydrogenases. This will be further augmented by allowing comparison of these data with those from the previously characterised GDH from *T. acidophilum*, the only other recombinantly expressed archaeal GDH to date.

The aim of this section of work is to characterise the recombinant *S. solfataricus* GDH in terms of its pH and temperature optima, its thermal stability at elevated temperatures, its substrate specificity with NAD^+, $NADP^+$ and a variety of mono- and disaccharides, and to determine its K_M and V_{MAX} values with a selection of substrates. Comparison of these data can then be made with the same properties obtained from the native *S. solfataricus* and other archaeal GDH enzymes. It is also hoped that analysis of the substrate specificity and kinetic data will help supply information on the nature of the active site in terms of its requirements for substrate binding.

6.2 MATERIALS AND METHODS

6.2.1 MATERIALS

Sodium acetate and glycine were obtained from Fisher Scientific, Loughbrough, Leicestershire, U.K. MOPS, EPPS, D-ribose, D-fucose, 2-deoxy-D-glucose, D-glucosamine, D-galactose, D-xylose, 6-deoxy-D-glucose, D-idose, D-allose, D-mannose, L-threose, maltose, lactose, sucrose, NAD^+ and $NADP^+$ were all obtained from Sigma Chemical Company, Poole, Dorset, U.K.

6.4.1 METHODS

All pH values quoted are those at the temperature at which the experiments were carried out and were in accordance with the $\delta pK_a/\delta t$ of that particular buffer (Dawson *et al.*, 1986).

6.2.2.1 GDH assay

As described in section 2.2.1 with the substitution of 100 mM HEPES pH 7, 20 mM $MgCl_2$ for 50 mM Tris/HCl pH 9, 20 mM $MgCl_2$.

6.2.2.2 pH profile

4×10^{-3} mg purified GDH protein (as described in chapter 5) was assayed for 2 min at 70°C in 100 mM of the following buffers: Acetate, MES, EPPS, Tris and Glycine at different pH values, but each within their respective pKa ranges. Care was taken to include overlapping pH values where one buffer was changed to the next. This would then rule out any ionic effects on the protein incurred by the buffer. All the assays also contained 20 mM $MgCl_2$, 0.5 mM $NADP^+$ and 2 mM glucose.

6.2.2.2 Temperature profile

Purified GDH protein was assayed for 1 min in 50 mM HEPES, pH 7, 20 mM $MgCl_2$, 0.5 mM $NADP^+$ and 5 mM glucose at the following temperatures: 50, 60, 70, 75, 80, 85, 90, 95 and 100°C. Another experiment to supplement this work was carried out at 70, 80 and 90°C, using exactly the same assay conditions, except that the substrates (0.5 mM $NADP^+$ and 5 mM glucose), 50 mM HEPES and 20 mM $MgCl_2$ were pre-incubated for 1 min prior to starting the assay by addition of GDH.

6.2.2.3 Thermal inactivation studies

GDH protein was incubated in 100 mM HEPES, pH 7, 20 mM $MgCl_2$, in the presence and absence of 50 mM glucose and 1 mM $NADP^+$ at a range of temperatures (70, 80, 90 and 95°C). Aliquots of 25 µl were take at specific time points and placed on ice. After all the samples had been taken,10 µl aliquots were assayed under normal assay conditions (70°C) to determine the residual activity of GDH.

6.2.2.4 Determination of Substrate specificity

Purified GDH protein was assayed for 2 min in 100 mM Tris/HCl, pH 8.5 at 70°C with 20 mM $MgCl_2$ and 2 mM of either of the following sugars: D-ribose, D-fucose, 2-deoxy-D-glucose, D-glucosamine, D-galactose, D-xylose, 6-deoxy-D-glucose, D-idose, D-allose, D-mannose, L-threose, maltose, lactose, sucrose. Assays were carried out in the presence of either 2.5 mM NAD^+ or 0.5 mM $NADP^+$.

6.2.2.5 Determination of K_m and V_{MAX} values

Purified recombinant GDH protein was assayed at 70°C in 100 mM HEPES, pH 7, containing 20 mM $MgCl_2$ with either of the following substrates: 25 mM D-glucose, 2 or 4 mM D-xylose, 10 mM D-galactose, 5 mM D-fucose, and 2 mM 6-deoxy-D-glucose. Assays were carried out in the presence of either 10 mM NAD^+ or 5 mM $NADP^+$.

Purified native GDH protein was assayed at 70°C, in 100 mM HEPES, pH 7 containing 20 mM $MgCl_2$ with either of the following substrates: 25 mM D-glucose and 2 or 4 mM D-xylose. Assays were carried out in the presence of either 10 mM NAD^+ or 5 mM $NADP^+$.

6.3 RESULTS

6.3.1 pH PROFILE

A variety of buffers covering a range of pH values as described in 6.2.2.1 were employed in order to determine the optimum pH value of GDH. The results shown in Fig 6.1 reveal that GDH possesses optimal activity at pH 10.5. However, raising this pH to 11 causes a dramatic decrease in GDH activity and precipitation of the protein, resulting in an asymmetrically shaped pH profile.

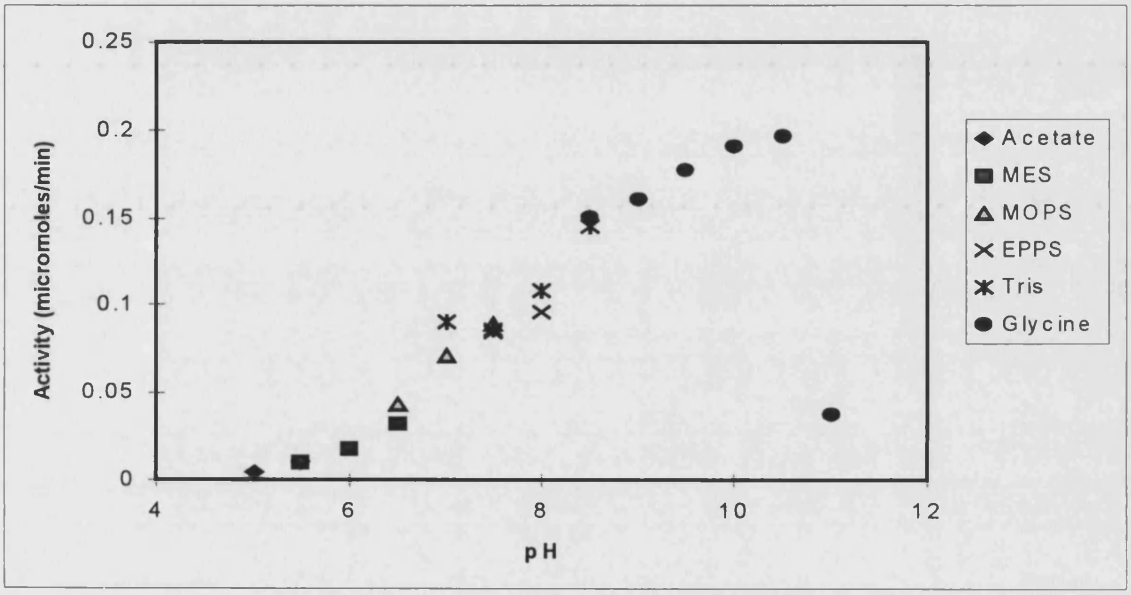


Fig. 6.1 **Activity of *S. solfataricus* GDH over a range of pH values in a variety of buffers.**

6.3.2 TEMPERATURE PROFILE

HEPES buffer, pH 7, was employed to assay GDH in order to determine its optimal temperature in accordance with thermal inactivation studies carried out using the GDH from *T. acidophilum*. Results revealed that GDH possessed optimal activity at 90°C (Fig. 6.2). However, an Arrhenius plot of these data (Fig. 6.3) revealed that inactivation of GDH, and consequently deviation from the linear relationship, began to occur above 70°C. Therefore, in order to deduce whether the observed deviation from linearity may be caused by cofactor instability, as reduced and oxidized forms of nicotinamide cofactors are both subject to decomposition in aqueous

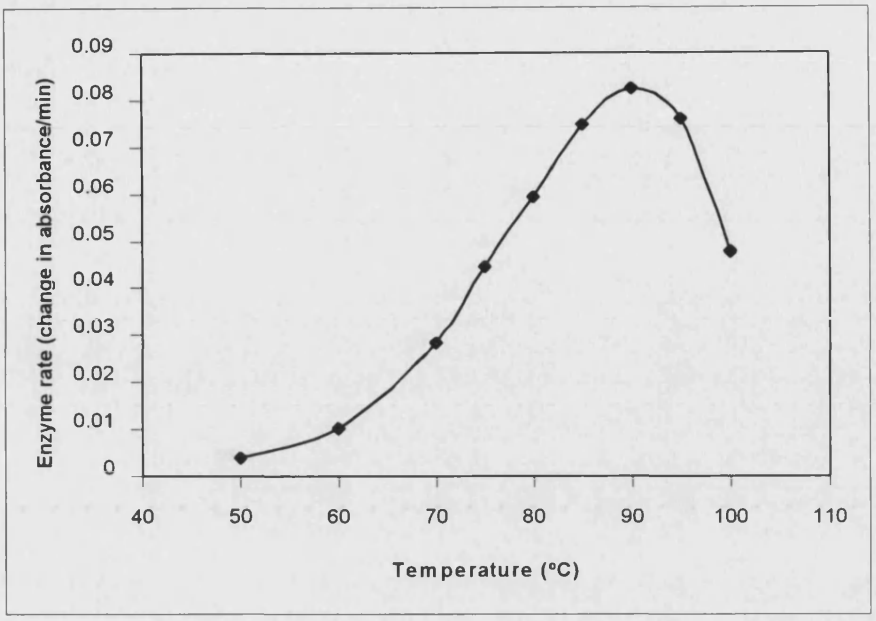


Fig. 6.2 **GDH activity at pH 7 on variation of temperature (°C).**

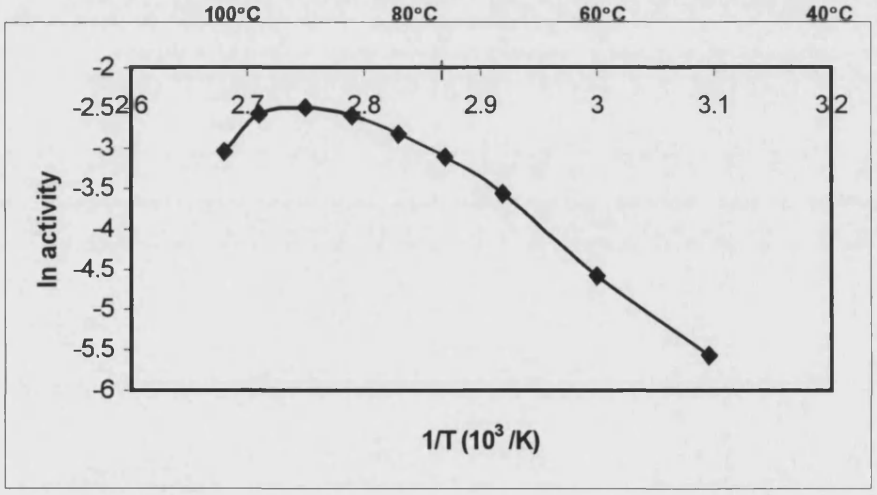


Fig. 6.3 **Arrhenius plot of GDH activity on varying temperature (°K).**

Exponential relationship between GDH actvity and the reciprocal temperature highlighted by the tangent to the curve.

solution, which is enhanced at elevated temperatures (Walsh *et al.*, 1983; Robb *et al.*, 1992; Daniel and Danson, 1994), supplementary experiments were carried out as described in 6.2.2.2. However, these showed that after 1 min at 90°C and 95°C and a subsequent assay under normal conditions

(at 70°C) the substrates were found to be only minimally degraded (Table. 6.1).

Temperature (°C)	% GDH activity remaining
70	100
90	95
95	90

Table. 6.1 The effect on GDH activity in a standard assay following a 1 min pre-incubation of $NADP^+$ at various temperatures.

All the values are expressed as a percentage of the activity found at 70°C with no incubation period.

6.3.3 THERMAL INACTIVATION

The thermal inactivation studies carried out at 70°C were with either GDH only or with GDH plus 50 mM glucose or 1 mM $NADP^+$. The studies at 80, 90 and 95°C were with GDH only or with GDH plus 50 mM glucose, as described in 6.2.2.3.

The studies carried out at 70°C revealed that GDH retained 100 % activity after 24 h, but this was severely reduced in the presence of substrates. The presence of 50 mM glucose resulted in 75 % residual activity after 24 h and the presence of 1 mM $NADP^+$ caused GDH to retain only 33 % of its activity after 24 h (Fig. 6.4). Studies carried out at 80°C revealed that GDH was only able to retain 54% of its total activity after 2.5 h. However, in this case the presence of 50 mM glucose had no effect on the rate of inactivation (Fig. 6.5). Studies carried out at 90°C resulted in 66 and 58 % residual activities after 1 h in the presence of 0 mM and 50 mM glucose respectively (Fig.6.6). Studies carried out at 95°C resulted in residual activites of 32 and 36 % after 1 h in the presence and absence of glucose respectively. These data differ from the others slightly, in that the presence of the glucose actually causes a minor increase in the residual activity of GDH at this elevated temperature (Fig. 6.7)

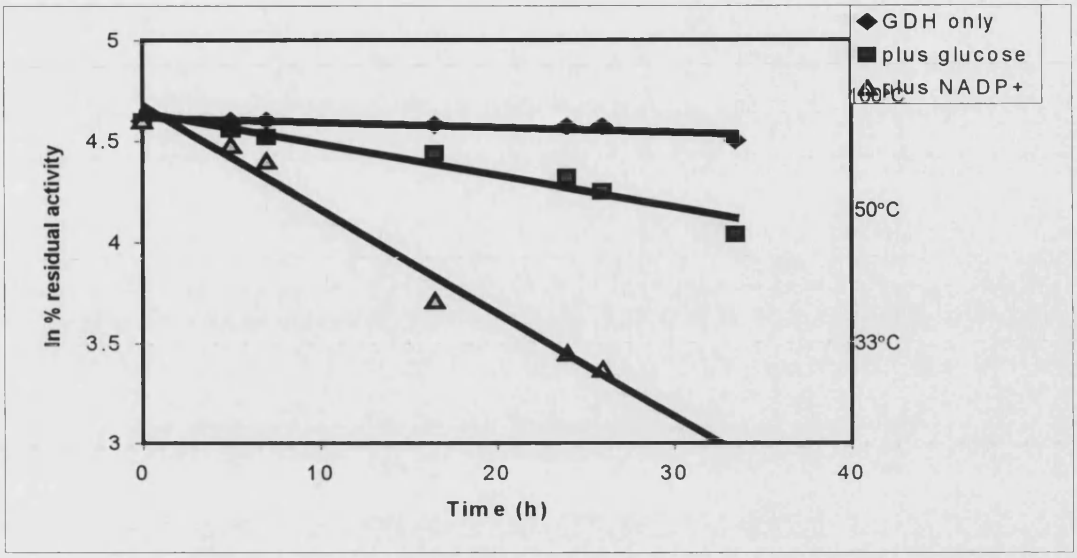


Fig. 6.4 **Thermal inactivation of GDH at 70°C in the presence and absence of glucose.**

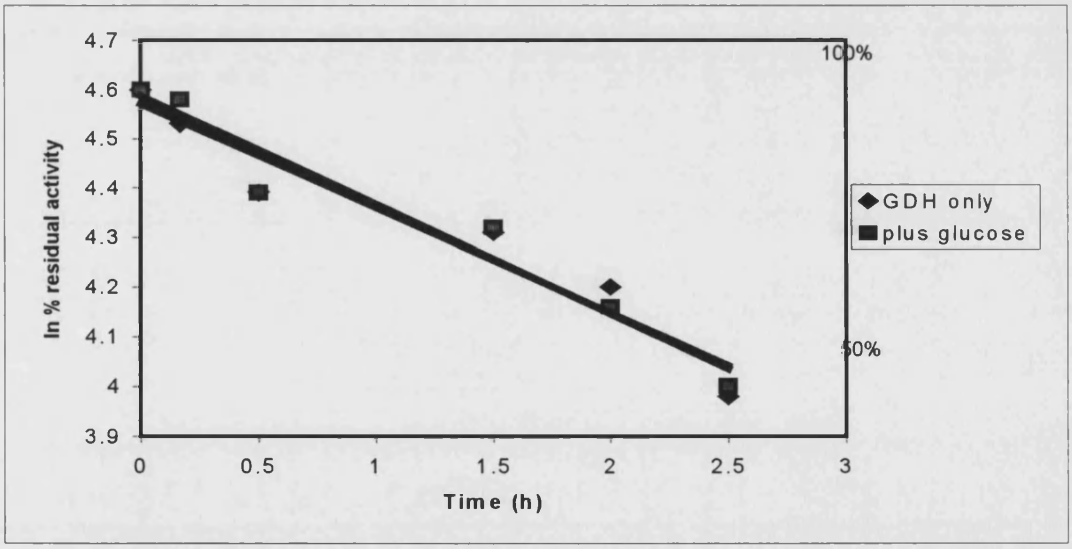


Fig. 6.5 **Thermal inactivation of GDH at 80°C in the presence and absence of glucose**

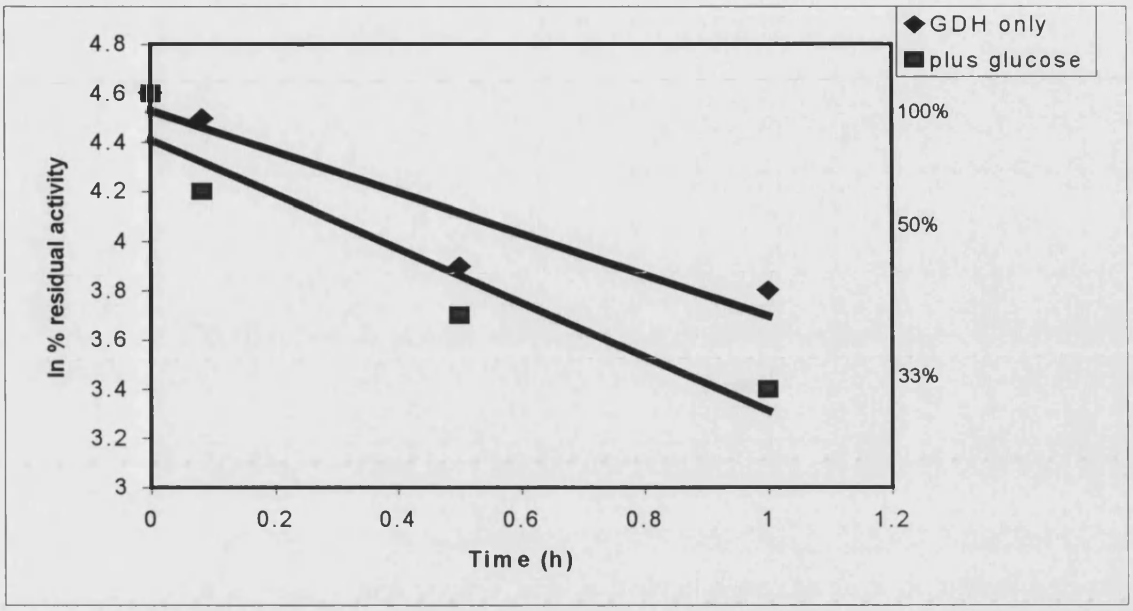


Fig. 6.6 **Thermal inactivation of GDH at 90°C in the presence and absence of glucose.**

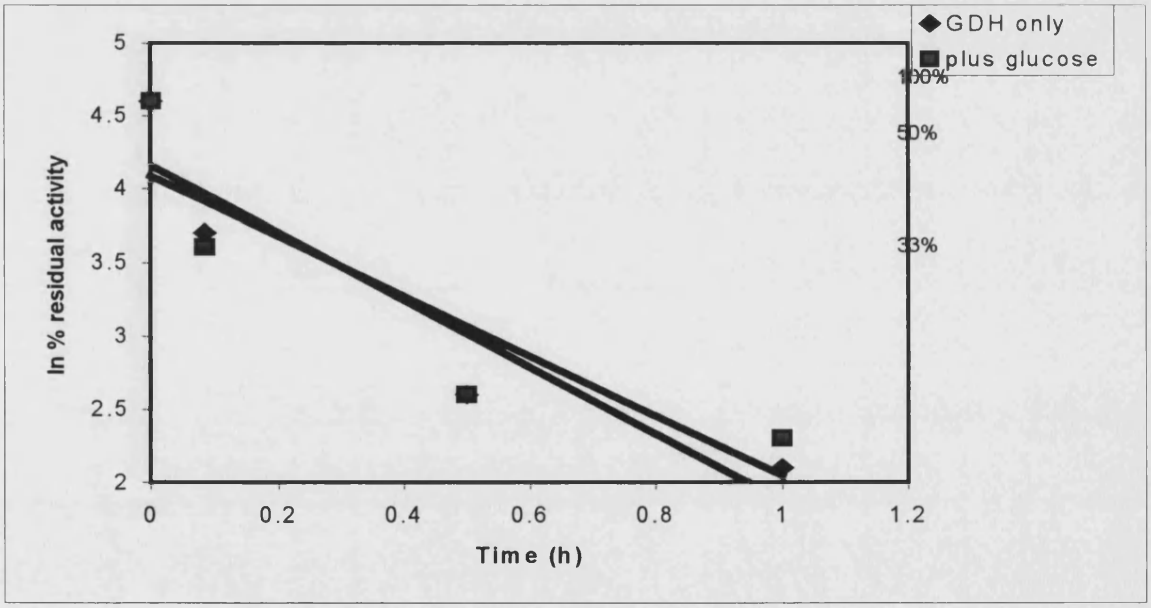


Fig. 6.7 **Thermal inactivation of GDH at 95°C in the presence and absence of glucose.**

The thermal inactivation data are displayed in the form of an Arrhenius plot.

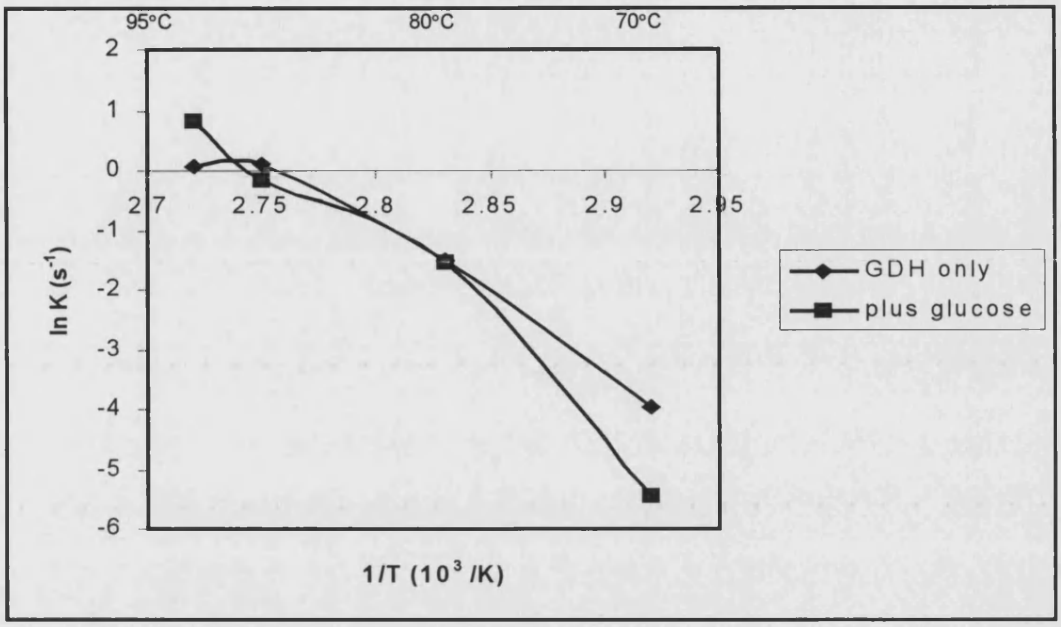


Fig. 6.8 **Arrhenius plot of the thermal inactivation of GDH in the presence and absence of 50 mM glucose.**

The data do not fit a straight line therefore it is not possible to calculate energy of activation (Ea) values for GDH in the presence and absence of glucose.

The plot reveals that, although the presence of glucose has an initial destabilising effect at 70°C, its continuing presence at increasing temperatures actually has an overall stabilising effect.

Similar thermal inactivation studies on the *T. acidophilum* GDH were carried out by Robinson, (1996). The data are given in Table 6.2 and show that the GDH from *S. solfataricus* is more thermostable than the *T. acidophilum* GDH, possessing almost 95 % of its total activity after 30 min at 80°C in comparison to a residual activity of only 50 % after 30 min at 78°C attained by the *T. acidophilum* GDH.

Species	% ACTIVITY REMAINING AFTER 30 MINUTES		
	Temperature (°C)	0 mM D-Glucose	50 mM D-Glucose
Sulfolobus solfataricus	80	94.5	85.6
Sulfolobus solfataricus	90	65.9	57.8
Sulfolobus solfataricus	95	32.0	36.0
Thermoplasma acidophilum	78	50.3	16.3

Table 6.2 **Residual activity of *S. solfataricus* and *T. acidophilum* GDH after 30 min incubation at various temperatures in the presence and absence of glucose.**

6.3.4 SUBSTRATE SPECIFICITY

Work carried out during the purification of the native GDH (chapter 3) revealed that GDH is able to oxidise D-xylose as well as D-glucose. Therefore, a more extensive survey of potential substrates for GDH was carried out. This involved trials with several monosaccharides including some epimers of D-glucose, and some disaccharides, which will indicate which groups on the D-aldose compounds are necessary for catalysis by comparing their relative oxidation rates. (Table 6.3).

Of all the sugars tested (Table 6.3), the most significant levels of activity were achieved with D-fucose, D-galactose, D-xylose, 6-deoxy-D-glucose and of course D-glucose. However, the levels of activity attained i.e. the rate of cofactor reduction, were not equal between the two cofactors. With the exception of D-glucose and 6-deoxy-D-glucose, which show higher percentage activities with $NADP^+$, all the other sugars with significant activity show preferential rates with NAD^+ as the cofactor. Activity determinations with C-4 and C-6 epimers of D-glucose showed that altering the configuration of C-4 group for instance D-glucose → D-galactose lowers the GDH activity obtained with NAD^+ by approximately 30%, to leave approximately 50% activity and lowers $NADP^+$ reduction by 90%, down to merely 10% of the activity obtained with D-glucose. Also substitution of the C-6 hydroxyl group by a methyl group, as described by 6-deoxy-D-glucose, results in unequal reductions in the GDH activities obtained with NAD^+ and $NADP^+$. Yet inclusion of both these changes in the form of D-fucose achieves the same activities as those obtained with D-galactose and is not additive i.e. is not lowered even further. Alternatively, a C-2 epimer of D-glucose, i.e. D-

mannose, gave no evidence of cofactor reduction with either NAD^+ or $NADP^+$.

Comparison of these data with substrate specificity data of other archaeal GDHs (*T. acidophilum*, *H. mediterannei* and *T. tenax*) (Table 6.4) reveals that they all appear to possess optimal activities with D-glucose and $NADP^+$, as indicated by the 100% relative activity rates. However, thereafter the individual enzymes have differing specificities. With the exception of the commonality of their preferences for glucose and $NADP^+$ and the high percentage rates achieved with D-xylose and NAD^+ for *S. solfataricus* GDH and *H. mediterranei* GDH, no other correlation could be found between the relative rates of *S. solfataricus* GDH and the other archaeal GDHs. However, none of the archaeal GDHs have been as extensively tested as the GDH from *S. solfataricus*.

	RELATIVE RATE (%)	
CARBOHYDRATE	NAD^+	$NADP^+$
D-Glucose	83	100
D-Ribose	<1	<1
D-Fucose	51	10
2-Deoxy-D-glucose	2.0	<1
D-Glucosamine	9.0	2.0
D-Galactose	53	12
D-Xylose	72	8.0
6-Deoxy-D-glucose	49	57
D-Idose	3.0	1.0
D-Allose	0	0
D-Mannose	0	0
L-Threose	0	0
Maltose	1.0	<1
Lactose	<1	1.0
Sucrose	1.0	1.0

Table 6.3 **Substrate specificity of *S. solfataricus* GDH.**

All compounds were at 2 mM concentration; the $NADP^+$-dependent D-glucose oxidation rate was taken as 100%. NAD^+ was tested at 2.5 mM concentration and $NADP^+$ was tested at 0.5 mM concentration.

SUBSTRATE	RELATIVE RATE (%)							
	S. solfataricus GDH		*T. acidophilum* GDH		*H. mediterranei*		*T. tenax* GDH	
	NAD^+ (2.5 mM)	$NADP^+$ (0.5 mM)	NAD^+ (10 mM)	$NADP^+$ (0.8 mM)	NAD^+ (6 mM)	$NADP^+$ (0.12 mM)	NAD^+ (7-50 mM)	$NADP^+$ (0.3-20 mM)
D-Glucose	83	100	13	100	26	100	√	100
D-Fucose	51.2	9.6	ND	16	3	18	ND	ND
D-Galactose	53.2	11.8	7	56	1	13	ND	0
D-Xylose	71.9	8.4	0	8	79	88	ND	√
D-Glucosamine	9.4	1.9	0	1	5	6	ND	ND
6-Deoxy-D-glucose	49.2	56.6	0.5	10	ND	ND	ND	ND
2-Deoxy-D-glucose	1.8	0.5	0	1	ND	ND	ND	ND
D-Ribose	0.62	0.1	0	0	-	3	ND	0
D-Idose	2.6	1.2	0	0	ND	ND	ND	ND
D-Mannose	0	0	0	0.3	7	7	ND	ND
D-Allose	0	0	0	7	ND	ND	ND	ND
D-Glulose	ND	ND	0	8	ND	ND	ND	ND
D-Altrose	ND	ND	0	0	ND	ND	ND	ND
L-Threose	0	0	ND	ND	ND	ND	ND	ND
D-Arabinose	ND	ND	ND	ND	ND	ND	ND	0
Fructose 6-phosphate	ND	ND	ND	ND	ND	ND	ND	0
Glucose 6-phosphate	ND	ND	ND	ND	ND	ND	ND	0
Sorbitol	ND	ND	ND	0	ND	ND	ND	0
Ethanol	ND	ND	ND	ND	ND	ND	ND	0
Erythritol	ND	ND	ND	ND	ND	ND	ND	0
Lactose	0.57	0.83	ND	4	ND	ND	ND	0
Maltose	1.2	0.33	ND	ND	ND	ND	ND	ND
Sucrose	0.82	0.79	ND	ND	ND	ND	ND	ND

Table 6.4 **Relative rates of sugar oxidation achieved with known archaeal GDHs.**

Concentrations of carbohydrate compounds tested were: 2 , 50, 100 and 10 mM with *S. solafataricus*, *T. acidophilum* (Smith, 1989; Rossjohn, 1994), *H. mediterranei* (Bonete *et al.*, 1996) and *T. tenax* GDH (Siebers *et al.*,1997), respectively. For each enzyme, the rate achieved with D-Glucose and $NADP^+$ was assigned a value of 100%. √ indicates evidence of utilization despite the absence of relative % rates. ND = not determined.

6.3.5 DETERMINATION OF K_M AND V_{MAX} VALUES

Kinetic analyses were carried out as described in 6.2.2.5 and K_M AND V_{MAX} data were determined from the direct linear plots of Eisenthal and Cornish-Bowden (1974) using Enzpack 3. K_{cat} /K_M values for GDH were calculated for a variety of sugars.

SUBSTRATE	K_M (μM)	V_{MAX} (μmole/min/mg)	K_{CAT} (s^{-1})	K_{CAT} /K_M ($s^{-1}mM^{-1}$)
NAD^+ (10 mM)				
Glucose	1500 (±50)	110 (±5)	75	50
Xylose	250 (±5)	90 (±10)	60	240
Galactose	570 (±10)	90 (±1)	65	110
Fucose	340 (±10)	40 (±1)	25	75
6-deoxy-D-glucose	80 (±5)	25 (±1)	20	220
$NADP^+$ (5 mM)				
Glucose	1300 (±50)	70 (±2)	50	40
Xylose	180 (±10)	65 (±3)	45	240
Galactose	440 (±10)	55 (±1)	40	85
Fucose	250 (±10)	30 (±1)	20	85
6-deoxy-D-glucose	80 (±5)	20 (±1)	15	190

Table 6.5a Kinetic data of sugars for GDH with NAD^+ and $NADP^+$ as co-substrates.

	K_M (μM)	
Co-substrates	NAD^+	$NADP^+$
Glucose	900 (± 25)	50 (±2.0)
Xylose	300 (±15)	25 (±1.5)
Galactose	550 (±30)	30 (±1.0)
Fucose	200 (±10)	20 (±0.8)
6-Deoxy-D-glucose	150 (±10)	80 (±10)

Table 6.5b Kinetic data of cofactors for GDH with a variety of sugar co-substrates.

Concentrations of carbohydrates tested were: glucose, 25 mM; xylose, 2 and 4 mM with NAD^+ and $NADP^+$ respectively; galactose, 10 mM; fucose, 5 mM; and 6-deoxy-D-glucose, 2 mM.

All the K_M values of sugars and cofactors obtained generally decrease in value amongst the sugars in the following order: D-Glucose, D-Galactose, D-Xylose, D-Fucose and 6-deoxy-D-glucose. The V_{MAX} and k_{CAT} values tend to follow the same pattern, with D-Glucose possessing the highest value and 6-deoxy-D-glucose the lowest. The K_{CAT}/K_M values obtained indicate that GDH is able to catalyse the oxidation of D-Xylose and 6-deoxy-D-glucose with the highest efficiency, using both cofactors. In fact the lowest catalytic efficiency was found with D-Glucose, also in the presence of both cofactors.

In order to ensure that no major changes in the catalytic function of GDH on expression in a mesophilic host it was also necessary to obtain the kinetic data concerning the cofactors and sugars with the native GDH. The resulting K_M and V_{MAX} values were compared and revealed to be very similar (Table 6.6a & b).

(a)

	***S. solfataricus* GDH (recombinant)**		***S. solfataricus* GDH (native)**	
Substrates and co-substrates*	**K_M (μM)**	**V_{MAX} (μmole/min/mg)**	**K_M (μM)**	**V_{MAX} (μmole/min/mg)**
NAD^{+}* (10 mM)				
Glucose	1500 (±50)	110 (±5)	650 (±25)	110 (±5)
Xylose	250 (±5)	90 (±10)	250 (±15)	70 (±5)
$NADP^{+}$* (5 mM)				
Glucose	1300 (±50)	70 (±2)	1200 (±50)	100 (±5)
Xylose	200 (±10)	65 (±3)	170 (±5)	70 (±5)

(b)

	***S. solfataricus* GDH (recombinant)**		***S. solfataricus* GDH (native)**	
Substrates and co-substrates*	**K_M (μM)**	**V_{MAX} (μmole/min/mg)**	**K_M (μM)**	**V_{MAX} (μmole/min/mg)**
Glucose*(25 mM)				
Xylose* (2 mM)				
NAD^{+}	900 (±25)	ND	500 (±15)	180 (±5)
NAD^{+}	300 (±15)	ND	600 (±90)	190 (±2)
Glucose*(25 mM)				
Xylose* (4 mM)				
$NADP^{+}$	50 (±0.05)	ND	55 (±1)	125 (±30)
$NADP^{+}$	25 (±0.05)	ND	35 (±2.5)	120 (±5)

Table 6.6a&b Comparison of the kinetic data obtained for the native and recombinant GDH from *S. solfataricus*. Co-substrates are indicated by *.

Comparison of the kinetic data obtained for *S. solfataricus* GDH was also made with those obtained for the *T. acidophilum* GDH (although were not possible for the V_{MAX} data). However, the K_M values for the *S. solfataricus* GDH were found to be considerably (5 – 60 fold) less than those of the *T. acidophilum* GDH.

	***S. solfataricus* GDH (native)**		***T. acidophilum* GDH (native)**	
SUBSTRATE	**K_M (μM)**	**V_{MAX} (μmole/min/mg)**	**K_M (μM)**	**V_{MAX} (μmole/min/mg)**
NAD^+ (cosubstrate)				
Glucose	650 (±25)	113 (±5)	20,000*	ND
$NADP^+$ (cosubstrate)				
Glucose	1200 (±50)	104 (±5)	9500*	ND
Glucose (cosubstrate)				
NAD^+	500 (±15)	ND	>30,000	>1.31 (±0.12)
$NADP^+$	55 (±1)	ND	300*	457*

Table 6.7 Comparison of the kinetic data obtained from *S. solfataricus* GDH and *T. acidophilum* GDH.

The concentrations of the co-substrates used were as follows: for *S. solfataricus* GDH - $NADP^+$ 5 mM, NAD^+ 10 mM and D-glucose 25 mM and for *T. acidophilum* GDH - $NADP^+$ 0.8 mM, NAD^+ 10 mM and D-glucose 50 mM as carried out by Smith, (1989) and * denotes concentrations of co-substrates employed by Rossjohn, (1994) for D-glucose, $NADP^+$ and NAD^+ were 0.5 - 8K_M.

6.4 DISCUSSION

6.4.1 pH PROFILE

The graph in Fig. 6.1 shows that optimal activity was obtained at pH 10.5 in 50 mM glycine buffer with 20 mM $MgCl_2$. However, it was not possible to go beyond this maximum and observe a decrease of activity due to a further increase in pH, as this caused the protein to precipitate out of solution. The increased pH allows the *S. solfataricus* GDH to function at its optimal activity through a variety of methods firstly presumably by changing the conformation and/or the ionization status of the enzyme and reactants. However, it is presently impossible to predict accurately the ionisation behaviour of any one residue or the titration curve of a total protein. This is because the

ionisation of each group on a protein is affected by its environment, which is determined by the protein, the solvent and by the ionisation of other groups on the protein.

Increasing the pH also corresponds to a decrease in proton concentration, which would effectively alter the position of equilibrium of the reaction to favour the release of protons from glucose, also resulting in the formation of gluconolactone. In addition, the alkaline conditions facilitate the spontaneous hydrolysis of gluconolactone to form D-gluconic acid and the increased stability of NADPH (Wong & Whitesides, 1981; Wu *et al.*, 1986) and therefore leads to elevated absorbance recordings (Fig. 6.1). However, increasing the pH above pH 10.5 is actually detrimental to the integrity of the GDH protein. This is due to non-specific repulsions that arise when a protein is highly charged, for example at extremes of pH. These repulsion events are classical electrostatic effects and are one of the ways in which electrostatic interactions can affect protein stability, the other being specific charge interactions, for example through ion-pairing (Dill, 1990). The square of the net charge is traditionally held to be a determinant of electrostatic free energy (Tanford, 1961). Hence, no electrostatic contribution to protein stability is expected near the isoelectric point (Tanford, 1968), at which the charges are evenly distributed across the protein surface. In the case of *S. solfataricus* GDH the isoelectric point was determined to be pH 7 (see chapter 4). However, on increasing the acidity or basicity of the solution, the net charge on the native protein is increased, which therefore results in an increase in charge repulsion and will destabilise the folded protein because the charge density on the folded molecule is greater than on the unfolded molecule.

However, despite the possible positive effects of raising the pH value, on the direction of the GDH reaction, no conclusions can be drawn as to the effects, of raising the pH of the environment on GDH activity. This is due to the non-saturating conditions of the substrates, carried out in this experiment and these results can only serve as an indication as to the optimal conditions for assaying GDH, i.e. if not all the GDH has bound substrate, the equilibrium of the GDH reaction to favour the oxidation of glucose can not be determined.

6.4.2 TEMPERATURE PROFILE

The variation of GDH activity with temperature (Fig. 6.2), indicates a deviation from the

exponential relationship at higher (and lower) temperatures. The deviation is such that the GDH activity achieved at 95°C is actually 30 % lower than that achieved in the absence of thermal inactivation. The temperature profile plot (Fig. 6.2) and Arrhenius plot (Fig. 6.3) both reveal the occurrence of inactivation of GDH above 85°C and 75°C respectively. However, reference to the supplementary experiments to deduce whether the observed inactivation was caused by cofactor instability (Table 6.1), revealed that following 1 min incubations of $NADP^+$ at temperatures of 70, 90 and 95°C, the respective GDH activities in standard assays were: 100, 95, 90 %. Therefore, the temperature-dependent degradation of this cofactor after just 1 min appears to be only minimal. This suggests that the decrease in activity shown by GDH above 75°C is a result of its inactivation and not dependent on the substrate stability. However, reference to the thermal inactivation data for GDH (6.3.2) reveals that in fact after 1 min at 95°C, the calculated residual activity is 97 %. This also does not correlate with the 30 % decrease in activity observed at 95°C. Consequently, if the decrease in activity seen in 6.3.2 is not a result of cofactor instability or thermal inactivation of the enzyme, then presumably the elevated temperatures are causing a partial unfolding of GDH that is reversible. This reversible conformational change may act by placing GDH in a lower state of activity.

6.4.3 THERMAL INACTIVATION

Thermostable proteins such as the enzyme GDH used in this study have been isolated from thermophiles and hyperthermophiles that have adapted to environments with temperatures from 50°C to the current upper limit of life, 113°C (Blochl *et al.*, 1997). These are temperatures that would denature most mesophilic proteins. Their thermostability is thought to be provided by the inclusion of a few significant stabilising interactions resulting in reduced conformational flexibility and increased resilience (see chapter 1, section 1.3.2). In addition to such stability at high temperatures, thermophilic enzymes also possess great resistance to proteolysis (Raia *et al.*, 1995) and detergent effects (Colombo *et al.*, 1992). However, in addition to the intrinsic stability of the proteins, in some hyperthermophilic Archaea, this stability has been aided *in vivo* and *in vitro* by the presence of high concentrations of intracellular metabolites (Hensel, 1993; Danson and Hough, 1998). Yet the *S. solfataricus* GDH was found to exhibit decreased thermal stability in the presence

of substrates as seen in 6.3.3. The presence of substrates and other low-molecular-mass compounds have been known to confer structure-stabilising effects (Jaenicke, 1991). They may participate to decrease the rate of unfolding by maintaining the structural conformation of the active site, yet it appears to be having the opposite effect here. However, it is necessary to look at the effect of temperature on the stability of the substrates themselves. If the substrates possess half-lives considerably less than 24 h, then the break down products of these compounds could indeed be acting as inhibitors. $NADP^+$ and NAD^+ are known to be very unstable at elevated temperatures, possessing half-lives of only minutes at 95°C (Walsh *et al.*, 1983; Adams *et al.*, 1992; Daniel and Danson, 1995). However, at 95°C the presence of glucose appears to provide a slight stabilising effect. Therefore, presumably the inhibitory products show the same properties as the substrates in providing a scaffold for the protein to wrap around and maintain its conformation.

The Arrhenius plot shown in (Fig. 6.8), is not strictly linear as expected, either in the presence or absence of glucose. The co-incubation of glucose with GDH at 70°C shows the only deviation from the linearity of the plot and has a destabilising effect on GDH, but the rate constant was maintained at higher temperatures. However, with GDH only, above 95°C, the rate constant of unfolding is no longer maintained. Production of non-linear Arrhenius plots of the rates of either unfolding or refolding or both, are thought to depend on the difference in heat capacity of the folded and unfolded states (Creighton, 1990). Therefore, the presence of metabolite is clearly decreasing the propensity towards the transition to the unfolded state

The thermal inactivation studies (Table 6.2) showed that *S. solfataricus* GDH was much more thermostable than the *T. acidophilum* enzyme. The *S. solfataricus* GDH appeared also to be less affected by the presence of glucose than the *T. acidophilum* GDH. Therefore, more detailed comparisons of these enzymes are required in order to yield another series of homologous enzymes that exist over a range of different growth temperatures in which to investigate thermostability. Examples of these series have been achieved with citrate synthase (Russel *et al.*, 1997; Russel *et al.*, 1994) and glutamate dehydrogenase (Yip *et al.*, 1995) (see chapter 1, section 1.3.2).

Although some degree of substrate inhibition was observed with certain substrates, these

occurrences were ignored for the purpose of this study and all the K_M and V_{MAX} data were obtained from direct linear plots. The increased thermostability and lower level of substrate inhibition at higher temperatures shown by *S. solfataricus* GDH were important for assuring the increased suitability of this enzyme over the *T. acidophilum* GDH for use in the hydrogen production process (chapter 7).

6.4.4 SUBSTRATE SPECIFICITY

The range of aldoses (4 - 6 carbons in length) tested in the presence of NAD^+ or $NADP^+$ are shown in Fig. 6.9 using Haworth projections in order to show them in their pyranose forms, which are the assumed forms of the sugar substrates (Schachter *et al.*, 1969). This will facilitate the observation of the relevant groups about the carbon chain atoms that are important in catalysis. However, chair and boat forms of these structures are actually more realistic than the Haworth projections, as carbon-carbon bond angles are then correctly portrayed.

Substantial levels of oxidation were observed with the following monosaccharides: D-glucose, D-fucose, D-glucosamine, D-galactose, D-xylose and 6-deoxy-D-glucose. In all cases, apart from D-glucose and 6-deoxy-D-glucose, the oxidation rate was higher in the presence of NAD^+. The *S. solfataricus* GDH does actually possess a nucleotide-binding motif, GXGXXG (Fig. 4.8), which would denote NAD^+ specificity and consequently suggest an NAD^+ preference. However, this is masked by the absence of other requirements for NAD^+ specificity, such as a negatively charged residue at the end of βB strand (John *et al.*, 1994), which in the *S. solfataricus* enzyme is an uncharged residue (Asn211).

Alterations in the configuration of groups at C-2 - C-6 show that GDH shows substrate specificity for non-phosphorylated D-aldoses with a glucose specific stereo-configuration at C-2, C-3 and C-4 (Fig. 6.10). Rotating the configuration of groups about the C-2 (D-mannose) causes a complete abolition of GDH activity. Yet substitution of the hydroxyl group at the C-2 of D-Glucose with an amino group, though still capable of undergoing oxidation, results in a large decrease in the rate of sugar oxidation to 10% and 2% with NAD^+ and $NADP^+$, respectively.

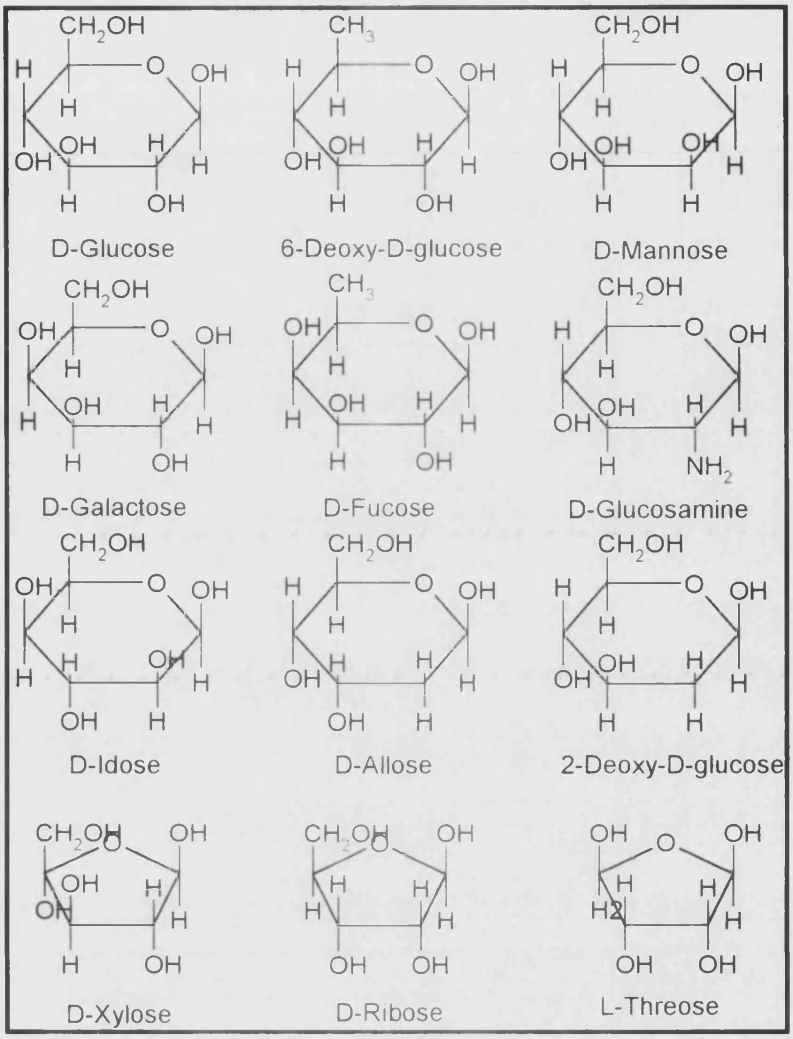


Fig 6.9 **Pyranose and furanose structures of monosaccharides used as potential substrates of GDH enzyme. All aldose sugars are depicted using Haworth projections**

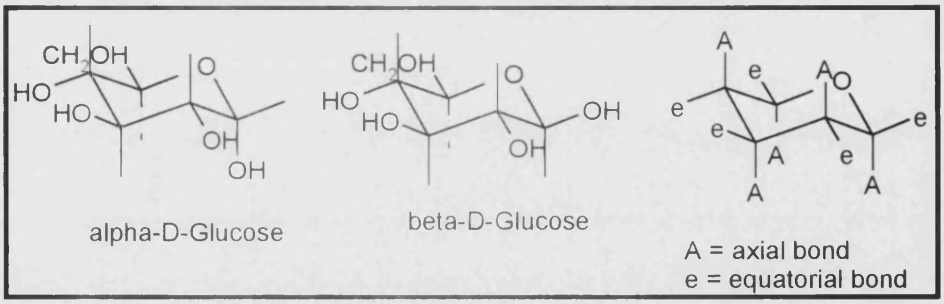


Fig. 6.10 **Diagram to show the generally preferred, chair forms of α and β-D-Glucose. The substituent groups about each carbon group are labelled "A" (axial) and "e" (equatorial).**

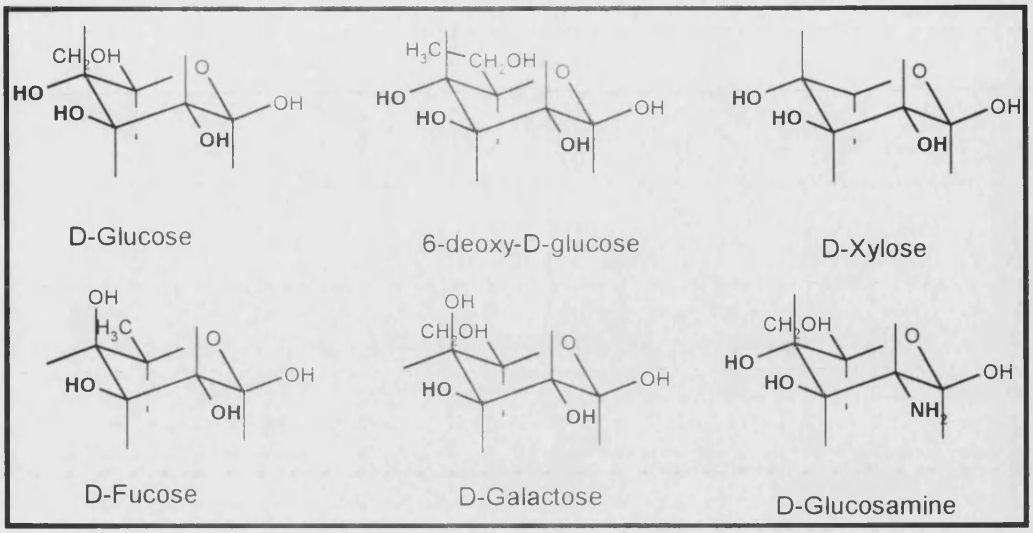


Fig. 6.11 **Depiction of the relevant aldoses substituents necessary for GDH activity. Equatorial nucleophilic substituents are shown in bold.**

Additionally, acceptance of the substrates D-xylose (a aldopentose sugar), 6-deoxy-D-glucose and D-fucose, albeit with lower oxidation rates, shows the diminished importance of the configuration of the C-6 groups. The depiction of these substrates, together with D-glucose, D-galactose and D-glucosamine as pyranose-rings in chair conformations (Fig. 6.11) highlights the commonality of all possessing equatorial non-H substituents at C-2, C-3 and C-4. That is with the exception of D-galactose and D-fucose (6-deoxy-D-galactose) that are C-4 epimers of D-glucose. In fact β-D-glucose is the only D-aldohexose that can simultaneously have all five non-H substituents in the equatorial position. This is due to the fact that large substituents generally prefer to be in an equatorial location due to a decreased incidence of steric hindrance with other groups and this may explain why glucose is the most abundant naturally occurring monosaccharide. This hypothesis concerning the requirement of equatorially positioned non-H substituents at C-2, C-3 and C-4, is in accordance with previous work carried out on other archaeal GDHs (Bonete *et al.*, 1996; Siebers *et al.*, 1997) including that on the GDH from *S. solfataricus* by Giardina *et al.* (1986) and additionally with several other previously described pyridine-nucleotide-dependent aldose dehydrogenases from bacterial and mammalian liver sources (Strecker and Korkes, 1952; Lee and Dobrogosz, 1965; Metzger *et al.*, 1964).

However, conflicting substrate specificity data were achieved with both GDHs isolated from *S. solfataricus* (this work and Giardina *et al.*, 1986) (Table 6.8), although it is possible that *S. solfataricus* may possess more than one aldose dehydrogenase. This is further augmented by the differences in molecular weights, amino acid compositions and kinetic data obtained for both these GDHs.

	RELATIVE RATE (%)			
	GDH in this work		Giardina *et al.* GDH	
CARBOHYDRATE	NAD^+	$NADP^+$	NAD^+	$NADP^+$
D-Glucose	83	100	100	9
D-Ribose	0.62	0.1	0	4
D-Fucose	51.2	9.6	ND	ND
2-Deoxy-D-glucose	1.8	0.5	12	25
D-Glucosamine	9.4	1.9	26	5
D-Galactose	53.2	11.8	0	15
D-Xylose	71.9	8.4	26	28
6-Deoxy-D-glucose	49.2	56.6	66	9
D-Idose	2.6	1.2	65	12
D-Allose	0	0	8	13
D-Mannose	0	0	0	4
L-Threose	0	0	ND	ND
D-Altrose	ND	ND	5	12
D-Glulose	ND	ND	8	12
Maltose	1.2	0.33	ND	ND
Lactose	0.57	0.83	ND	ND
Sucrose	0.82	0.79	ND	ND

Table. 6.8 Different percentage activities achieved with a variety of aldoses using *S. solfataricus* GDH.

(a) GDH isolated in this work and (b) GDH isolated by Giardina *et al.*, 1986.

Studies with the GDH isolated from the work showed that L-Threose, a four-carbon chain (aldotetrose) sugar with a levorotatory form, (which does not exist in nature), did not serve as a substrate, despite it possessing the same stereoconfiguration about C-2 and C-3 as D-Glucose, if present in a furanose conformation. However, its incompatibility would suggest the requirement of aldopyranoses in the dextrorotatory conformation in order to achieve aldose oxidation with *S. solfataricus* GDH. They also showed that limited oxidation was attained with disaccharides, although this is most probably due to their contamination with glucose monomers.

Unequivalent percentage decreases in activity were seen with the GDH-activities obtained using NAD^+ and $NADP^+$. Presumably the difference in activities relates to the phosphate content of

the coenzymes.

Comparison of the substrate specificity data obtained for *S. solfataricus* GDH with other archaeal GDHs (Table. 6.4) reveals a commonality in the preferential use of D-Glucose and $NADP^+$. This suggests an evolutionary relationship between these species that evolved to oxidise aldopyranose sugars (and possibly aldofuranose sugars, although this conformation is not very common) via direct hydrogen bonding side chains to the C-2, C-3 and C-4 hydroxyls. Perhaps divergence from a common ancestor or low stringency of glucose binding due to its universal abundance has led to the broad specificity range of these GDHs. This would enable the *S. solfataricus* organism to scavenge its surroundings for a variety of aldose-based energy sources. Significant levels of oxidation were obtained with D-Xylose, D-Galactose and D-Fucose for *T. acidophilum* and *H. mediterranei*. However, the limited substrate specificity testing of the other archaeal GDHs and the failure to test *S. solfataricus* GDH under saturating concentrations of sugars would affect any conclusions drawn from this work.

6.4.5 KM AND VMAX DATA

The kinetic data for the recombinant *S. solfataricus* GDH with the two cofactors and five aldose compounds (Tables 6.5 a and b), displayed a general decrease in the K_M values of the cofactors and sugars in approximately the following series: D-glucose>D-galactose>D-xylose>D-fucose>6-deoxy-D-glucose. The assumption that the rate of the chemical step, k_{CAT}, is slower than either k_1 or k_{-1}, under steady-state reaction conditions lead to the further assumption that K_M is therefore a measure of the affinity of E for S. Hence, the lower the value of K_M the more tightly the substrate is bound. Therefore, the aforementioned series of sugars appear to possess steadily higher binding affinities for GDH when placed in the reverse order.

K_M values are optimised by evolution so that they are effective for catalysis, i.e. are high enough to achieve proximity, but low enough that the substrates are not too tightly bound. The higher affinity of the C-4 and C-6 epimers of D-glucose (D-galactose, D-xylose, D-fucose and 6-deoxy-D-glucose) leads to the conclusion that the glucose-specific-configuration at the C-2 and C-3 position (but not the C-4 as suggested from the substrate specificity data in section 6.4.4) of D-glucose is preferred for recognition of sugars by *S. solfataricus* GDH and therefore suggests

specific interactions with hydrogen-bonding side chains of residues in the GDH active site. The configuration of groups about the remaining carbon atoms does not appear to be important for recognition and in fact, rotating or reducing their size leads to improved recognition and binding.

The K_M values exhibited by the cofactors are usually found to be low amongst enzymes. This is due to their high degree of specific interaction with residues in the active site. However, a noticeable reduction in K_M values is only shown with $NADP^+$, and therefore may imply a specific interaction between the ribose-phosphate group of $NADP^+$ and the enzyme.

The maximum velocity values (V_{MAX}) of sugar oxidation show an opposing change of values across the K_M series with D-glucose possessing the highest value and 6-deoxy-D-glucose the lowest. Therefore indicating that, in saturating concentrations of the various substrates tested, GDH is able to achieve the highest velocity of aldose oxidation with D-glucose.

The rate constant k_{CAT}/K_M is a measure of two features of enzymic catalysis, which includes substrate specificity and the catalytic efficiency of an enzyme. The k_{CAT}/K_M data obtained for the aldoses tested, does not follow the same pattern of values indicated by the above series. All reactions with D-xylose and 6-deoxy-D-glucose exhibited the highest catalytic efficiency. In addition closer inspection of the all the kinetic values for D-xylose, it was revealed to be a much more suitable substrate for *S. solfataricus* GDH than D-Glucose, in that it possesses a higher binding affinity (lower K_M), a comparable V_{MAX} value and a higher catalytic efficiency. However, accounting for the availability and the abundant universal supply of D-glucose, it may therefore not be necessary for *S. solfataricus* GDH to possess such a high affinity for it, although in the event of a depletion of this principle energy source, *S. solfataricus* GDH would require the ability to bind alternative aldoses with sufficient affinity in order to maintain the viability of the organism. Noting the origins and supply of D-xylose and the other substrates tested, shows that : D-xylose is actually a major sugar constituent of wood, D-galactose is widely found in nature in gums and fruit pectins, fucose in the form of L-fucose is a widely occurring polysaccharide component and 6-deoxy-D-glucose does not actually occur naturally. Yet the occurrence of these naturally appearing sugars (including D-glucose) in the hot sulfur springs from which *S. solfataricus* originates (De Rosa *et al.*, 1975; Zillig *et al.*, 1980) could be rather limited.

7 Chapter Seven

Hydrogen production

7.1 INTRODUCTION

7.1.1 WHY HYDROGEN ?

Hydrogen, a colourless, odourless, tasteless, flammable gas, is found at concentrations of about 0.0001% in the air and constitutes about 10% of our seas. Hydrogen is raising much interest in recent years, both politically and scientifically as the "superfuel of the future" (Lawler, 1995; Lepowski, 1995). This is due not only to its limitless supply, whereas fossil fuel reserves are limited, but also due to its ability to burn without generating any toxic by-products, where the only by-product of hydrogen fuel being water.

The potential of switching to a hydrogen economy is many fold. From a political point of view this would lead to less reliance on oil imports from the Middle East and result in countries becoming energy self-sufficient, thus changing the whole global economy. The environmental standpoint also looks promising in that the conversion to hydrogen energy would dispense with further oil exploration and so protect pristine wilderness areas. This in turn would also prevent the common occurrence of huge oil spills in transportation of exported oil, which kills marine life forms. Also, the cleanliness and efficiency of hydrogen would have a huge impact on the quality of air, thus decreasing the incidence of acid rain and greenhouse-gas emissions.

Hydrogen is produced by several methods, including steam/methane reforming (Wilson and Newall, 1970), dissociation of ammonia, and recovery from by-product streams from chemical manufacturing and petroleum reforming. Storage and transportation of hydrogen can be facilitated in a gaseous or cryogenic liquid state.

At present hydrogen is widely used in petroleum refining and chemical processes, metal processing operations, in the electronics industry and more famously as rocket fuel.

However, it appears that hydrogen's uses may be more far reaching as a major fuel source. Much investment and technology has already gone into developing hydrogen as an alternative transportation fuel (Appendix, 1-3). These transportation projects are currently being run as pilot operations, with the aim of developing into fuel cell powered public and private transport systems, with even a view to expanding into fuel cell driven submarines and fishing vessels (Leslie, 1997). This will lead to more long-term development and investment into the building of power plants, storage and manufacture of hydrogen and also to the establishment of fueling stations, the first of which have been launched in two major German cities (Appendix, 1 & 3). However, despite all the promise of this potential new "superfuel", much controversy is being made about the manufacture of hydrogen. This is due to the absence of a practical natural source of hydrogen. Therefore, it must be made by transforming some other energy source at a cost. Even electrolysis of sea water, a vast reserve of hydrogen, would require the use of electricity. Additionally, the manufacture of hydrogen from petroleum and methane eventually leads to the release of carbon dioxide, thus creating a cost to the environment. Consequently this has led to the discovery of hydrogen manufacture from renewable and non-costly and non-polluting sources.

7.1.2 ALTERNATIVE METHODS OF HYDROGEN PRODUCTION

(i) the conversion of municipal solid waste, automobile shredder residue and other plastic/ rubber wastes to hydrogen using a pyrolysis step followed by partial oxidation in a high-temperature, high-pressure gasifier (Wallman *et al.*, 1998).

(ii) the conversion of water to hydrogen without consuming conventional sources such as coal and electricity. This involves the use of light energy to convert water to hydrogen via photocatalytic semi-conductor particles, in the presence or absence of metals, electron-donors/acceptors and hole scavengers (Pedroni *et al.*, 1996; Ashokkumar, 1998).

(iii) the conversion of cellulosic waste to hydrogen via an *in vitro* enzymatic pathway (Woodward *et al.*, 1996; Woodward and Orr, 1998; Inoue *et al.*, 1999). This involves the conversion of potential glucose sources such as cellulose by cellulases, and plant sap (i.e. sucrose) by invertase and glucose isomerase to glucose (Woodward *et al.*,

1996) The glucose is then oxidized to glucono-δ-lactone and the cofactor, $NADP^+$ is simultaneously reduced. The presence of a pyridine-dependent-hydrogenase in this system (Egerer *et al.*, 1982; Bryant and Adams, 1989), causes the regeneration and recycling of $NAD(P)^+$ with the concomitant production of molecular hydrogen (Fig. 1.6).

This aim of this section of work was to determine if the GDH from *S. solfataricus* would improve the efficiency of this conversion system, through studies that would alter all the possible variables, namely enzyme, substrate and product concentration, temperature and pH.

7.2 MATERIALS AND METHODS

7.2.1 MATERIALS

7.2.1.1 Enzymes

Glucose dehydrogenase from *S. solfataricus* was expressed in *E. coli* from which the enzyme was purified using only a heat step (85°C, 30 min) and assayed as described in Chapter 5. The enzyme possessed a specific activity of 305 units/ml, where 1 unit is defined as the amount of enzyme required to produce 1 μmol NAD(P)H/min at 70°C, pH 7.0. Hydrogenase from *P. furiosus* was purified as described by Bryant and Adams, (1989) and assayed by measuring hydrogen evolution from reduced methyl viologen in 50 mM Tris/HCl, pH 8.0, containing 1 mM sodium diothionite at 80°C. Its catalytic activity was measured to be 519 μmoles/min/ml.

7.2.1.2 Chemicals

β-$NADP^+$, HEPES, MES, $MgCl_2$, spectinomycin, sodium azide, D-gluconic acid lactone, molecular sieve (calcium-alumino-silicate pellets, 3.2 mm) and glycine were all obtained from Sigma Chemical Company, St. Louis, M.O., U.S.A. β-D-Glucose was obtained from Calbiochem, San Diego, C.A., U.S.A.

7.2.1.3 Equipment

Hydrogen production was measured using a continuous flow system, as shown in Fig. 7.2, whose construction is described in detail by Greenbaum, (1984). This system was continuously purged with helium and calibrated with an inline electrolysis cell connected in tandem with the hydrogen detection system and according to Faraday's law of electrochemical equivalence. The hydrogen detection apparatus consisted of a combustible gas analyzer (Bio-Gas Detector Corporation, Okemos, MI, USA) consisting of a Figaro semiconductor tin oxide gas sensor. A voltage increase occurs across the sensor when a combustible gas comes into contact with the sensor surface due to a decrease in resistance; this was measured by a Keithley (Cleveland, OH, USA) Model 485 picoameter. Data were analyzed using ASYST Technologies, Inc., 4.0 Analysis Software (Rochester, N.Y., U.S.A).

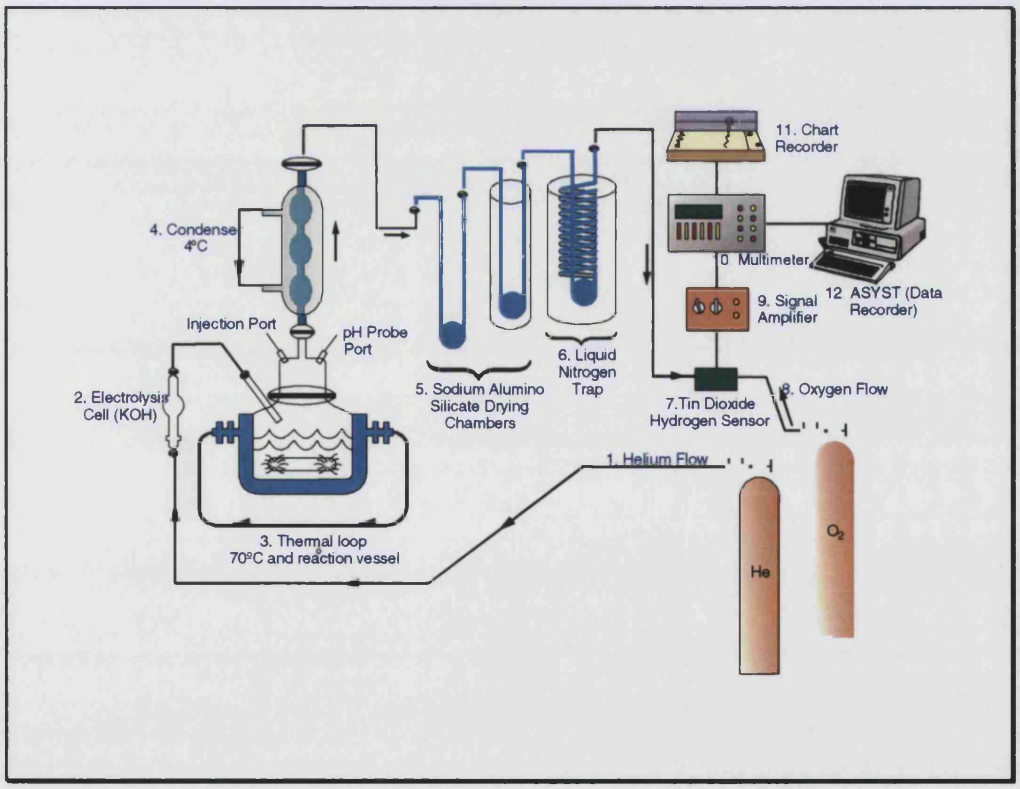


Fig. 7.1 Schematic depiction of the apparatus used to measure hydrogen production.

Diagram kindly provided by J. Getty, O.R.N.L., Oak Ridge, T.N., U.S.A.

7.2.2 METHODS

7.2.2.1 HYDROGEN PRODUCTION

On complete purging and temperature equilibration (40 - 70°C) of the system depicted in Fig. 7.1, the following constituents (given at their final concentrations) were added to the reaction vessel via syringes :

200 mM	HEPES/ MES/ Glycine buffers (pH 6-9)
20 mM	$MgCl_2$
0.1 mg	Spectinomycin
0.1 mg	Sodium azide
2.5 - 50 mM	Glucose
0.05 - 1.0 mM	$NADP^+$
2 Units (μmoles/min)	GDH (305 μmoles/min/ml stock)
6 - 15 Units	Hydrogenase (519 μmoles/min/ml stock)
FINAL VOLUME OF 2 ml	

The pH values of all the buffers were adjusted at room temperature so that, knowing their δpKa/δt coefficients, the desired pH values at the operating temperatures were achieved. The reaction was started with the addition of hydrogenase. The electrolysis cell (item 2 in Fig 7.1) usually employed in the calibration of the system, was also employed at the same position prior to the reaction vessel to act as a water reservoir due to the substantial level of evaporation encountered at these temperatures.

7.2.2.1.1 Effect of varying the hydrogenase concentration

The conditions were as described in section 7.2.2.1 with the variation in the concentration of hydrogenase added from 6 - 15 Units. The glucose was at a concentration of 20 mM, $NADP^+$ at 0.5 mM, HEPES at 200 mM, pH 7 and the amount of GDH was always maintained at 2 units. The reaction was maintained at a temperature of 70°C.

7.2.2.1.2 Effect of varying the pH

The conditions were as described in section 7.2.2.1 with the exception of the buffer type used. The reactions were carried out at pH 6, which employed 200 mM MES; at pH 7, which used 200 mM HEPES and at pH 8, using 200 mM Glycine. All were titrated with NaOH to values that accounted for the temperature dissociation constant of the buffers at 70°C. 20 mM glucose, 0.5 mM $NADP^+$, 2 units of GDH and 12 units of hydrogenase were present in all the reactions.

7.2.2.1.3 Effect of varying temperature

The conditions were as described in section 7.2.2.1 with 200 mM HEPES at pH 7, 2 units of GDH, 12 units of hydrogenase and initial concentrations of glucose at 20 mM and $NADP^+$ at 0.5 mM. However, each of the reactions was maintained at temperatures of 40°C, 50°C, 60°C and 70°C respectively.

7.2.2.1.4 Effect of varying NADP concentration

The conditions were as described in section 7.2.2.1 with an initial glucose concentration of 20 mM, 2 Units of GDH and 20 Units of hydrogenase. The buffer used was 200 mM HEPES at pH 7, and the $NADP^+$ concentration was varied from 0.05 - 1.0 mM. All reactions were carried out at 60°C.

7.2.2.1.5 Effect of varying glucose concentration

The conditions were as described in section 7.2.2.1 with an initial $NADP^+$ concentration of 0.25 mM, 2 Units of GDH and 20 Units of hydrogenase. The buffer used was 200 mM HEPES at pH 7 and the glucose concentration was varied from 2.5 - 50 mM. All reactions were carried out at 60°C.

7.2.2.2 EFFECT OF GLUCONIC ACID

GDH at a concentration of 0.5 μg/ml was incubated for 24 h in the presence of 0 - 50 mM gluconic acid and subsequently assayed under assay conditions mimicking those of the reaction vessel (100 mM HEPES, pH 7, 20 mM $MgCl_2$, 0.5 mM $NADP^+$ and 5 mM glucose, at 70°C over 2 min). It was also assayed (using the same conditions) in the presence of 0 - 100

mM gluconic acid. The gluconic acid in the form of D-gluconic acid lactone was pre-titrated to pH 7.0 prior to use.

7.3 RESULTS

7.3.1 EFFECT OF HYDROGENASE CONCENTRATION

Raising the concentration of hydrogenase in the reaction vessel led to an increase in the overall rate of hydrogen production from approximately 2500 to 8000 nmols/h (Fig. 7.2a). It also increased the overall stoichiometric yield of hydrogen produced to a maximum of 60% after 24 h (Fig. 7.2b).

7.3.2 EFFECT OF PH

Raising the initial pH value from pH 6 – pH 8 caused different behaviours in the rate of hydrogen production (Fig. 7.3a). The highest rate and yield was achieved with HEPES buffer at pH 7. An initial pH of 8 gave the lowest rate of hydrogen production but accomplished an overall yield of approximately 10% less than that achieved at pH 7. This was due to it maintaining a fairly substantial rate for longer (≈ 10h) as opposed to the very rapid rates carried over shorter periods (≈ 5h) seen in many of the other plots. The initial pH of 6 produced a higher rate of hydrogen production than that attained with pH 8 buffer, but it gave the lowest overall yield of only 22% (Fig. 7.3b).

7.3.3 EFFECT OF TEMPERATURE

Varying the temperature of the reaction over 24h also had different influences on the behaviour of the rate of hydrogen production (Fig. 7.4a). The highest rate was achieved at 70°C, followed by 60°C reaction. The plot produced by the reaction carried out at 50°C maintained a fairly substantial rate for over 10h. This was opposed to the virtual cessation in hydrogen production seen after 6 h at 70°C and 60°C. However, at 40°C, even though the rate was very low, (maximum rate approximately 2000 nmol/h), it was maintained over a period of over 40 h (not actually shown on the graph, Fig. 7.4a). Despite the large difference in rates of hydrogen production, the overall yields of hydrogen produced were very similar, with a mean value of 54% (Fig. 7.4b).

7.3.4 EFFECT OF $NADP^+$ CONCENTRATION

Varying the concentration of $NADP^+$ from 0.05 - 0.25 mM caused an increase in the rate and yield of hydrogen produced (Fig. 7.5a). However, raising the $NADP^+$ concentration further to 1 mM caused a large decline in the rate, which was not maintained and appeared to cease after only 3 h. Therefore, this only achieved an overall yield of 5%, in contrast to the 50% overall yield of hydrogen attained with 0.25 mM $NADP^+$ (Fig. 7.5b).

7.3.5 EFFECT OF GLUCOSE CONCENTRATION

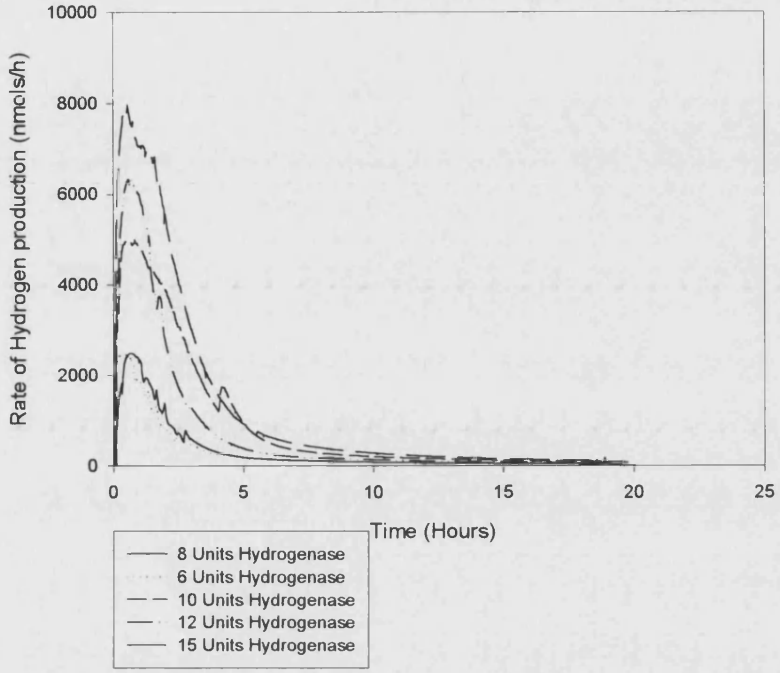
Varying the concentration of glucose from 2.5 - 10 mM caused a large increase in the rate and overall yield of hydrogen produced (Fig. 7.6a and b), in which stoichiometric yields were achieved with 10 mM glucose. However, a reaction containing 50 mM glucose failed to achieve the high rate of hydrogen production accomplished by the reaction containing 10 mM glucose and also only managed to attain 20% of its maximal yield.

All details of each experiment and their resulting yields in mmol are described and summarised in Table 7.1.

7.3.6 EFFECT OF GLUCONIC ACID CONCENTRATION

Incubation of GDH for 24 h with increasing concentrations of gluconic acid (pre-titrated to pH 7), caused a decrease in the activity of GDH. The presence of 50 mM gluconic acid caused a reduction in activity of approximately 99%. Confirmation of the occurrence of inactivation was received by the presence of increasing amounts of gluconic acid (also pre-titrated to pH 7) in standard assays (7.2.2) containing 5 mM glucose. The presence of 100 mM gluconic acid caused a 34% decrease in the activity of GDH over the duration of the assay (2 min).

(a)



(b)

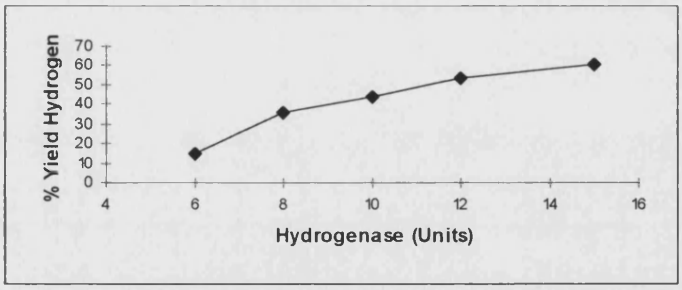
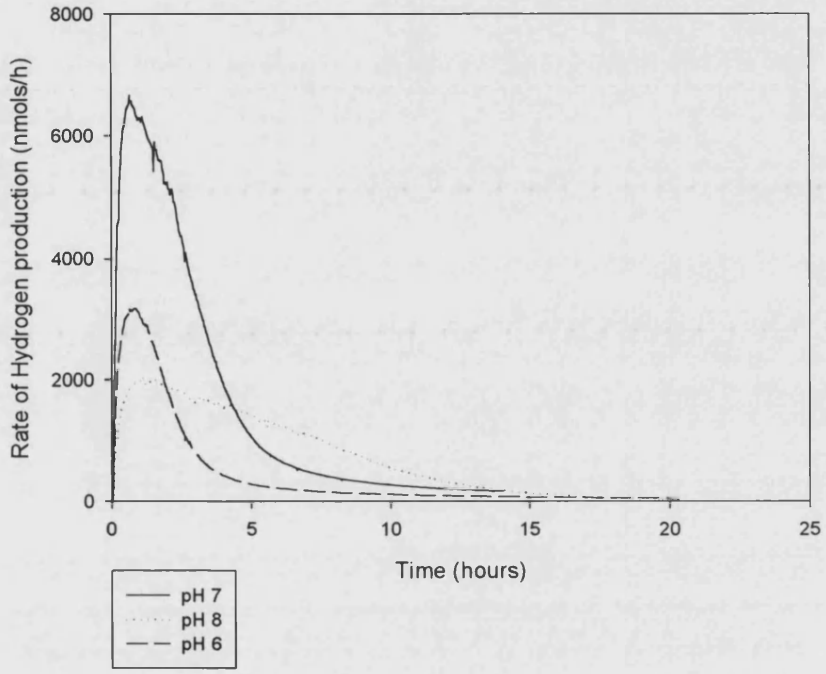


Fig. 7.2 The effect of hydrogenase concentration on hydrogen production.

(a) The rate of H_2 production (nmol/h) at different hydrogenase concentrations. (b) The overall % yields achieved with different concentrations of hydrogenase.

(a)



(b)

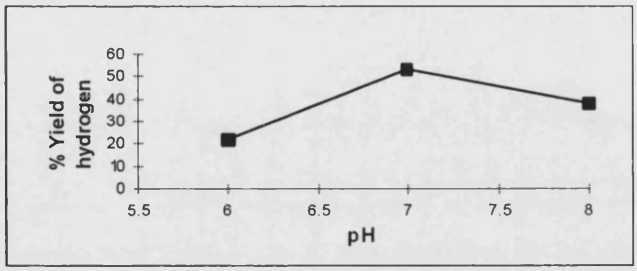
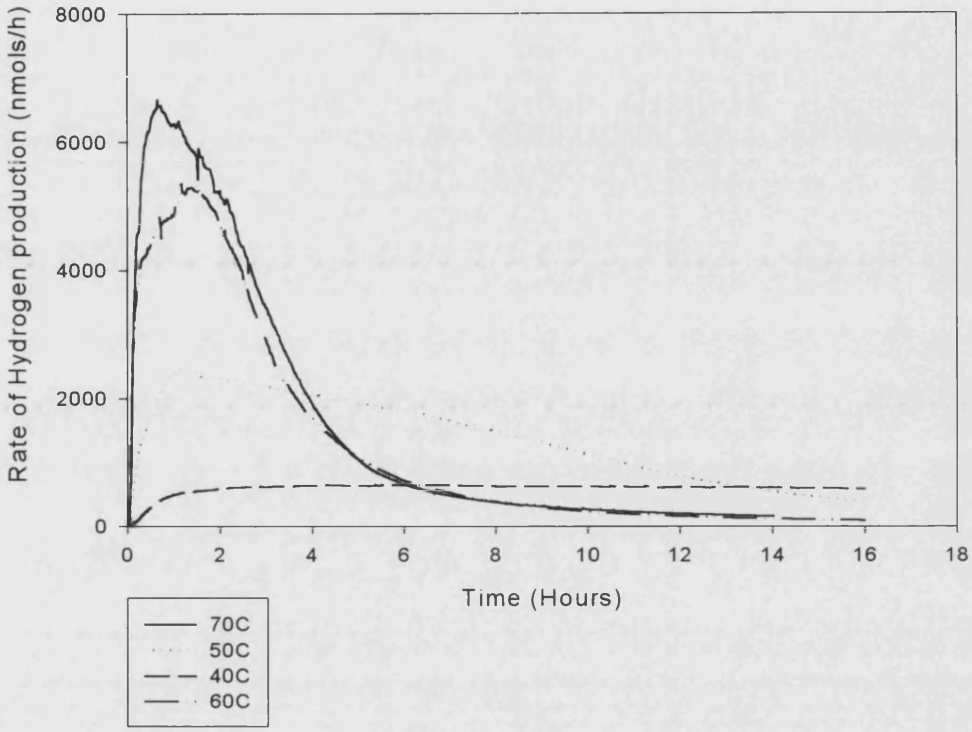


Fig. 7.3 The effect of pH on hydrogen production.

(a) The rate of H_2 production (nmol/h) at different pH values. (b) The overall % yields achieved at these pH values.

(a)



(b)

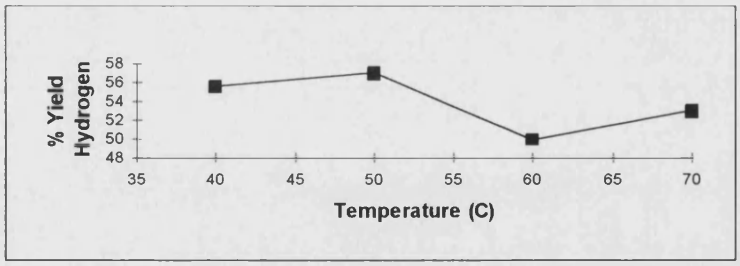
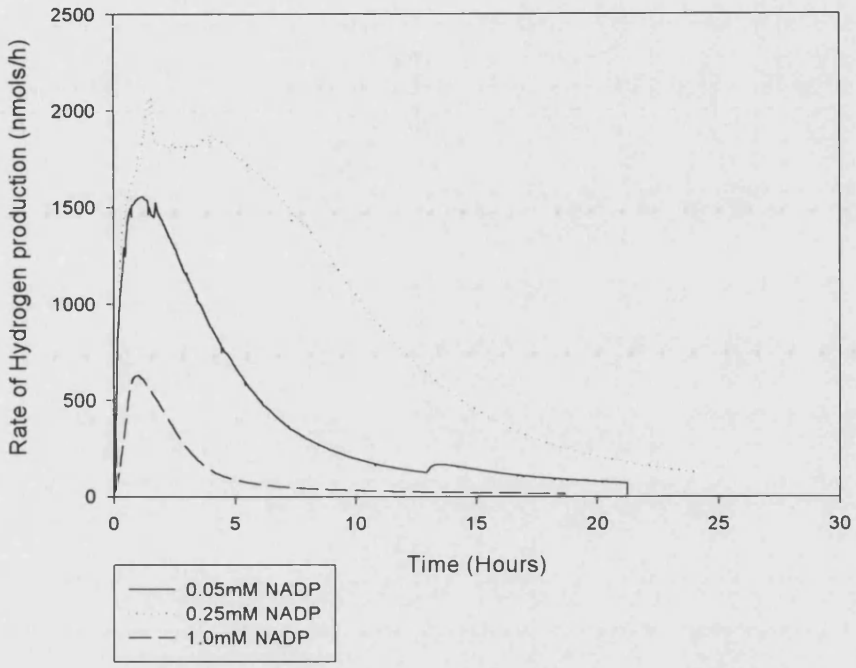


Fig. 7.4 The effect of temperature on hydrogen production.

(a) The rate of H_2 production (nmol/h) at different temperatures. (b) The overall % yields achieved at these temperatures

(a)



(b)

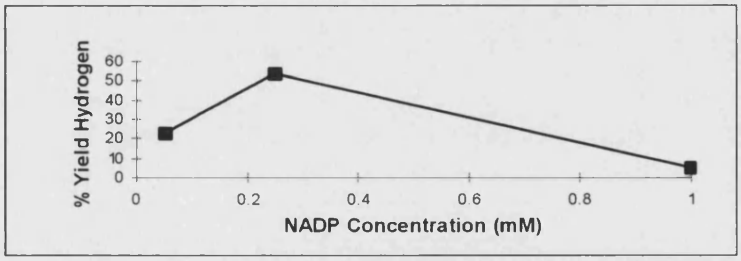
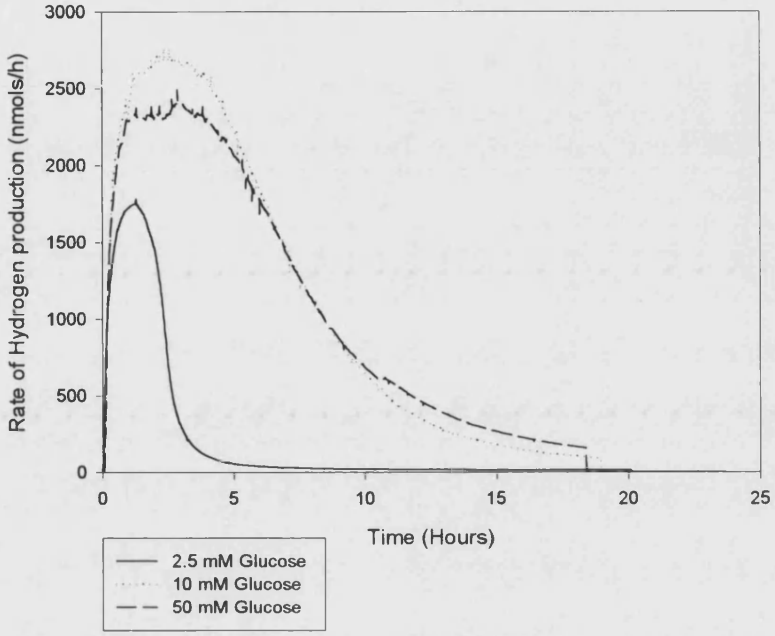


Fig. 7.5 Graphs showing the effect of $NADP^+$ concentration on hydrogen production.

(a) Graph showing the rates (nmols/h) produced at different NADP+ concentrations. (b) Graph showing the overall % yields achieved at these concentrations.

(a)



(b)

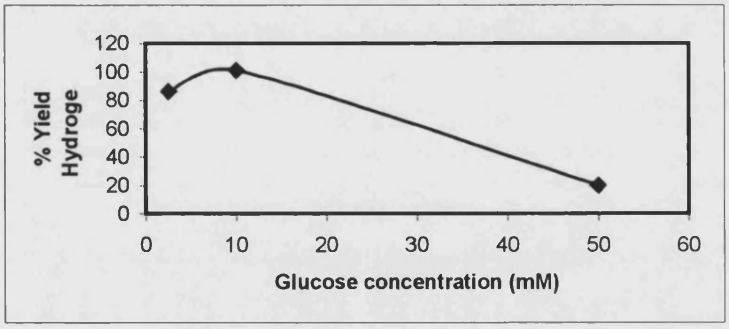
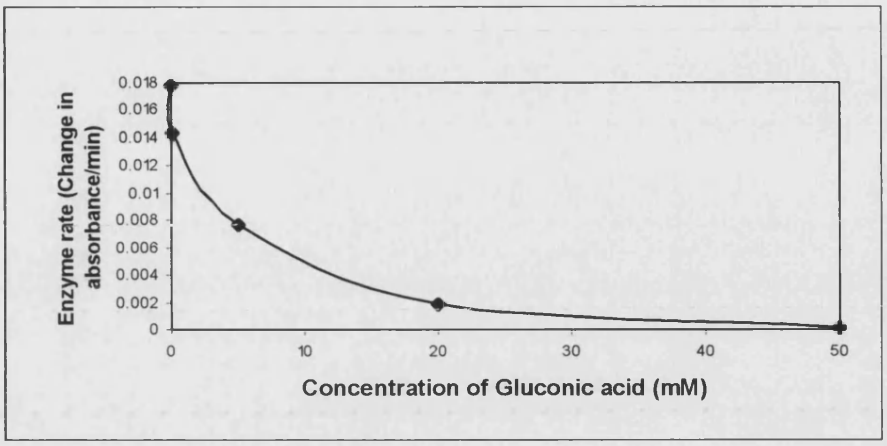


Fig. 7.6 The effect glucose of concentration on hydrogen production.

(a) The rate of H_2 production (nmol/h) at different glucose concentrations. (b) The overall % yields achieved at these concentrations.

(a)



(b)

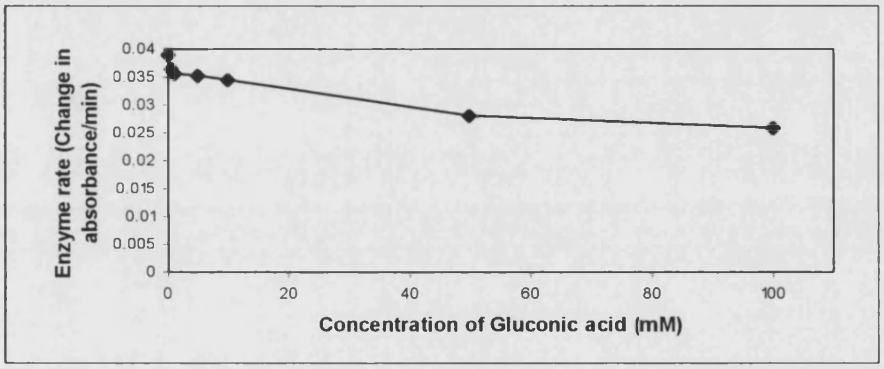


Fig. 7.7 The effect of gluconic acid on GDH activity.

(a) The effect on GDH activity of incubating GDH with increasing amounts of gluconic acid at 70°C over a 24h period. (b) The effect on GDH activity of the presence of increasing amounts of gluconic acid over a 1 min assay. Controls for both experiments employing no gluconic acid were carried out and are described as the first data point on each graph.

Variable	GDH concentration (units)	hydrogenase concentration (units)	temperature (°C)	pH	Glucose concentration (mM)	$NADP^+$ concentration (mm)	Yield of hydrogen (mmol)	% Yield of hydrogen
Hydrogenase	2	6	70	7	20	0.5	5.8×10^3	15
Hydrogenase	2	8	70	7	20	0.5	1.4×10^4	36
Hydrogenase	2	10	70	7	20	0.5	1.7×10^3	44
Hydrogenase	2	12	70	7	20	0.5	2.1×10^4	53
Hydrogenase	2	15	70	7	20	0.5	2.4×10^4	61
pH	2	12	70	6	20	0.5	8.7×10^3	22
pH	2	12	70	7	20	0.5	2.1×10^4	53
pH	2	12	70	8	20	0.5	1.5×10^4	38
Temperature	2	12	40	7	20	0.5	2.2×10^4	56
Temperature	2	12	50	7	20	0.5	2.3×10^4	57
Temperature	2	12	60	7	20	0.5	1.9×10^4	47
Temperature	2	12	70	7	20	0.5	2.1×10^4	53
NADP concentration	2	20	60	7	20	0.05	8.9×10^3	23
NADP concentration	2	20	60	7	20	0.25	2.2×10^4	54
NADP concentration	2	20	60	7	20	1.0	2.0×10^3	5
Glucose concentration	2	20	60	7	2.5	0.25	4.3×10^3	86
Glucose concentration	2	20	60	7	10	0.25	2.0×10^4	101
Glucose concentration	2	20	60	7	50	0.25	2.0×10^4	20

Table 7.1 Summary describing details of all variable experiments.

7.4 DISCUSSION

7.4.1 EFFECT OF HYDROGENASE CONCENTRATION

The rate and yield of hydrogen produced appears have a directionally proportional relationship with the hydrogenase concentration. The requirement for a larger ratio of hydrogenase to GDH could be due to one or a combination of many factors, including the thermostabilty of the hydrogenase, its catalytic efficiency (turnover), its suitability to the reaction conditions and/or its sensitivity to gluconic acid or the break down products of the cofactor. Firstly, in terms of its thermostablity, the hydrogenase has been found to possess a $t_{50\%}$ (a 50% residual activity) at 80°C of approximately 21 h (Bryant and Adams, 1989). In view of the fact that the optimal temperature of the hydrogenase is over 95°C in H_2-oxidizing activities (Fiala and Stetter, 1986) and 85°C with $NADP^+$ as a cofactor (Ma *et al.*, 1994), the running temperature of the reaction suggests that the hydrogenase is operating at an inferior capacity. The effectiveness of the electron shuttle between GDH and the hydrogenase may therefore be an important rate-determining factor. NADH and NADPH have been shown to be quite unstable, having half-lives of only a few minutes at neutral pH at 95°C (Walsh *et al.*, 1983; Robb *et al.*, 1992; Daniel and Danson, 1994). The stability of cofactor may also be a major determinant in the rate and yield of hydrogen evolution in this system. In support of this, addition of second aliquot of $NADP^+$ to this system (carried out at 50°C with GDH from *T. acidophilum*) after the cessation of hydrogen evolution, restored the rate of hydrogen production. Furthermore, this restoration was not dependent on the further addition of hydrogenase or GDH (Woodward and Orr, 1998).

Observation of the rate plots (Fig. 7.3a) reveals that addition of hydrogenase produces a higher maximal rate of hydrogen evolution, and it also maintains a substantial rate for longer. Since the thermostabilty of the hydrogenase is not in question and the GDH has been found to possess a $t_{50\%}$ of 24 h at 70°C (see chapter 6, section 6.3.3), this suggests that perhaps the greater amounts of hydrogenase might actually be protecting the cofactor by sequestering it to the active site and/or by maintaining it in a habitual electron shuttle between the GDH and the hydrogenase. However, previous studies have shown the rate of cofactor (NADH and NADPH) degradation to be unaffected by the presence of a dehydrogenase enzyme (Daniel and

Danson, 1994). The major breakdown products of the decomposition of these cofactors are thought to arise from cleavage of the nicotinamide-ribose linkage, as found with the destabilization of NAD^+ by spleen NADase (Zatman *et al.*, 1953). The resulting ADP-ribose and nicotinamide are found to be inhibitors of some dehydrogenases (Fawcett *et al.*, 1961; Lowry *et al.*, 1961; Wenz *et al.*, 1976) and studies have shown the breakdown products to be impervious to re-reduction by a hydrogenase and an appropriate co-substrate (Daniel and Danson, 1994). It is therefore possible that these end products may also contribute to inhibition of hydrogen production by occupying the active site of the hydrogenase or the GDH. Thermal inactivation experiments carried out with GDH at 70°C showed that a residual activity of 30% after 26 h in the presence of 1 mM NADP. This was opposed to a 100 % residual activity after 24 h with GDH that had been incubated in the absence of NADP (see chapter 6, section 6.3.3). The production of gluconic acid may also contribute to the amount of hydrogen evolved due to its contribution to the lowering of the pH of the reaction. Unlike hydrogenase (Ma *et al.*, 1993), GDH retains significant activity below pH 6.0 (see chapter 6, section 6.3.1) and, therefore, although glucose is capable of being oxidized to gluconic acid by GDH with the resulting fall in pH, the ability of hydrogenase to oxidize NADPH would be severly dimished below pH 6.0.

7.4.2 EFFECT OF pH

The pH optimum for the GDH reaction has been estimated to be pH 10.5 (see Chapter 6, Fig. 6.1) and the hydrogenase has an optimum of pH 8 for oxidizing reactions (the hydrogenase from *P. furiosus* is a "bifunctional" sulfhydrogenase possessing H_2-oxidizing and S^o-reduction activities that respond differently to pH, temperature and inhibitors (Ma *et al.*, 1993)). The achievement of such a low rate of hydrogen evolution at a pH of 8 is therefore somewhat surprising. This response was mirrored in the system containing GDH from *T. acidophilum* (Woodward and Orr, 1998); however, the explanation in this case was that the GDH possesses virtually no activity at pH 8. It can therefore only be hypothesized that although the enzymes are operating at or near their optimal pH values, this pH may not be favourable to the reaction system as a whole. The reduced forms of NAD^+ and $NADP^+$ are very stable in dilute alkali but labile in acid; the oxidized forms are stable in dilute acid but labile in base (Wong and Whitesides, 1981). At intermediate pH values reduced forms become unstable as

the pH is decreased and oxidized forms become more unstable as the pH is increased. The decomposition of NAD(P)H in pH 5 – 9 is mainly due to general acid catalysis (Wong and Whitesides, 1981; Oppenheimer and Kaplan, 1974; Woodward and Orr, 1998) and the major breakdown products of decomposition of these cofactors are ADP-ribose and nicotinamide, which are found to be inhibitors of some dehydrogenases (Fawcett *et al.*, 1961; Wenz *et al.*, 1976). However, the major pathway for the decomposition of $NAD(P)^+$ is by nucleophilic attack of the pyridine ring to form a 1,4-dihydropyridine structure (Johnson and Smith, 1976; Biellman *et al.*, 1979; Everse *et al.*, 1971; Chenault and Whitesides, 1987). Furthermore, several salts such as phosphate, acetate, NaCl, KCl are also shown to destabilise these coenzymes (Wong and Whitesides, 1981; Wu *et al.*, 1986; Rover *et al.*, 1998). Studies on a variety of unbuffered and buffered solutions have found the pH that caused the least destruction of either cofactor mixture ($NAD(P)^+$/NAD(P)H) to be pH 8.5 (Chenault and Whitesides, 1987). Therefore, the production of hydrogen in this system may benefit from pH values exceeding pH 8. However, without a facility to prevent the decrease in pH over time, the susceptibility of cofactor to instability at specific pH values will prevent the realisation of maximal hydrogen yields.

Additionally, it is not known whether the substrates are undergoing inactivation by other enzyme activities present in the reaction mixture (coupled to the hydrogenase or GDH due to the fact that they were not purified to homogeneity (7.2.1.1)), that favour an environment at pH 8 to pH 7, when accounting for the decrease in pH over the time of the reaction. Yet, observation of the rate plot for the pH 8 reaction actually reveals that, although the rate itself is not very substantial, it is actually maintained for a longer duration than the reactions carried out at pH 6 and pH 7. This is presumably due to the lower pH values brought about by gluconic acid production at those initial pH values. This would also suggest a reason for the sudden fall off in rate also shown by the reactions at pH 6 and pH 7, which occurs a lot earlier, without the complete conversion of glucose, at 3 h and 5 h respectively. The drop in pH could not only be detrimental to the stability NADPH but could also have an effect on the activity of the hydrogenase. Therefore, maintaining a constant pH in the reaction mixture could actually result in higher yields and allow the maintenance of high hydrogen evolution rate for longer periods.

7.4.3 EFFECT OF TEMPERATURE

The thermophilicity of the enzyme and the thermolability of the cofactor in this process leads to a paradoxical choice of reaction conditions. However, carrying out this reaction at different temperatures, actually resulted in similar overall % yields*. The plots at 60°C and 70°C exhibit the highest rates (Fig. 7.5a), due to the proximity of the enzymes to their optimal temperatures† . Yet these rates both decrease rapidly after approximately 5 h. This is not only due to gluconic acid production but also due to $NADP^+$/NADPH instabilty and declining substrate concentration. In contrast, the plots at 50°C and 40°C exhibit intermediary and very low rates, respectively, due to the remoteness of the enzymes from their optimal temperatures, thus reducing their catalytic capacities. Yet these rates are maintained over longer periods than those found at higher temperatures and is shown by the rate of hydrogen evolution at 40°C (Fig. 7.4a), which was virtually unchanged after 40 h. Therefore, in spite of low rates of turnover, the cofactor is more stable at these temperatures and can therefore persist, presumably until completion of the reaction ie. 100% conversion of glucose to hydrogen).

7.4.4 EFFECT OF $NADP^+$ CONCENTRATION

The decline in the rate of hydrogen evolution after the maximum rate has been reached can be argued using several explanations, namely the effect of sustained high temperatures on the stability of $NADP^+$/NADPH, the effect of increasing acidity due to gluconic acid production and also the decline in substrate concentration over time during the experiment. Previous work has shown that further addition of substrate, when hydrogen production has virtually ceased, results in a restoration of the maximum rate of hydrogen evolution (Woodward *et al.*, 1996). However, studies on the effect of initial substrate ($NADP^+$) concentration on rates and yields of hydrogen production does not follow this pattern. On increasing the initial concentration of $NADP^+$ from 0.05 to 2.5 mM, a large increase in hydrogen evolution is observed (Fig. 7.5), but at concentrations of 1.0 mM, the maximum rate was lower than both the previous concentrations. This observation can be explained by substrate inhibition due to the saturating levels of end products of $NADP^+$ breakdown, which would occur

* Although it must be noted that the reaction carried out at 40°C continued for twice the allotted time of the other reactions.

† Temperature optima, GDH = 90°C and hydrogenase = 85°C.

more readily if the $NADP^+$ molecules are not involved in the "electron shuttle" due to their excess numbers.

7.4.5 EFFECT OF GLUCOSE CONCENTRATION

Evidence of substrate inhibition is also observed on varying initial concentrations of glucose, where raising the initial concentration of glucose from 2.5 to 10 mM causes a large increase in the rate and yield of hydrogen evolved. Yet on raising this concentration further to 50 mM glucose (which is approximately 50 times the K_M of *S. solfataricus* GDH for glucose with NADP, K_M = 1.3mM), the maximum rate, instead of being five times greater, is comparable to that observed with a 10 mM initial glucose concentration. This equates to yields of 5 μmol hydrogen in both cases. However, their overall % yields are very different with a 100% yield attained with 10 mM glucose and merely 20% attained with 50 mM glucose. This could be explained by the change in the reaction conditions with time. The plot profiles for 10 mM and 50 mM initial glucose concentrations are almost superimposable and both rates plots fall off considerably after 10 h. At this time there would be a substantial amount of gluconic acid present in the reaction mixture, which would affect the evolution of hydrogen for reasons stated in section 7.4.1. Therefore, unless the initial reaction conditions can be maintained, complete conversion of amounts of glucose exceeding 50 mM is impossible.

7.4.6 EFFECT OF GLUCONIC ACID CONCENTRATION

The effects of gluconic acid production on the reaction process have already been proposed through its pH lowering effects. Yet it seems that it also causes the inactivation of GDH in some way, as the maintenance of constant pH by carrying out a pre-titration of gluconic acid negates the possibility that this inactivation effect is simply due to increasing acidity. The difference in the linear and cyclic structures of gluconic acid and the actual product of glucose oxidation, glucono-1,5-lactone, respectively (see Chapter 1, Fig. 1.9), and the absence of reversibility of the loss of activity on dilution, mean that the observed decrease in GDH activity can not be due to inhibition and is probably the result of inactivation.

7.4.7 SUMMARY

Overall, there was a slight improvement in the production of hydrogen by employing the *S. solfataricus* GDH in this system as opposed to the *T. acidophilum* GDH. Stoichiometric yields of 100% were possible with 10 mM glucose, 0.25 mM $NADP^+$, 2 units *S. solfataricus* GDH, 12 units hydrogenase, 200 mM HEPES, pH 7, at 70°C. But the *T. acidophilum* GDH was only able to achieve stoichiometric yields with the following conditions, 5 mM glucose, 0.5 mM $NADP^+$, 13 units of GDH, 26 units of hydrogenase, in 50 mM sodium phosphate buffer, pH 7, at 50°C. However, the requirement for constant conditions in this reaction system are absolute and the complete stoichiometric conversion of greater amounts of glucose to hydrogen may be possible if the pH could be maintained at pH 8.5 (Chenault and Whitesides, 1987) and removal of gluconic acid could be facilitated. Carrying out these requirements would necessitate the use of immobilization. The enzymes could therefore be immobilized on some type of support and the reactants can be constantly pumped in with the simultaneous removal of product. This would eliminate substrate inhibition, a drop in pH and gluconic acid build up by constant limited supply of the formers and constant removal of the latter. Another option is to actually immobilize the cofactor. Preliminary tests have shown this to have a large positive influence on the rates and yields of hydrogen evolution, when compared to equal amounts of unimmobilized cofactor (H.M. O'Neill, personal communication).

8 Chapter Eight:

Modelling & crystallisation

8.1 INTRODUCTION

8.1.1 WHY CRYSTALLISE PROTEINS?

Knowledge of protein macromolecular structure has many potential uses, including the ability to view the effects of structural and functional modifications on proteins, created through mutagenic and genetic engineering studies, to define ligand interactions with a view to the rational design of drugs and pharmacological agents, and to gain further insights into the regulation, mechanism, stabilising factors and detailed function of proteins.

There are presently numerous physical-chemical approaches that yield information regarding macromolecular structure. However, few of these approaches are capable of determining the structures of proteins to an atomic resolution. A structure can be visualised only if light with a wavelength comparable to its dimensions is used. Visible light, which is electromagnetic radiation, possesses wavelengths of 400 – 700 nm. As bonded atoms are only about 0.15 nm or 1.5 Å apart, visible light cannot produce an image of individual atoms in protein molecules. Therefore, visualisation requires the use of X-rays, electrons and neutrons, which possess the correct dimensions.

X-rays and neutrons are scattered by a protein in solution, but they cannot be focussed to reconstruct an image of a single molecule, as seen with visible light using a light microscope; consequently, only limited data concerning protein dimensions can be obtained. However use of these types of radiation in X-ray or neutron diffraction analysis of single crystals, and nuclear magnetic resonance (NMR) analysis of small proteins in solution, allow the elucidation of the three-dimensional structure of proteins to high resolution.

Advantages of opting for X-ray diffraction analysis of single crystals of proteins over NMR is that is not limited by the protein dimensions.

8.1.2 CRYSTALLISING PROTEINS

Elucidation of a protein structure by x-ray or neutron diffraction analysis requires three-dimensional crystals of the protein. They must be well ordered, with virtually every protein molecule held in a specific position in the three-dimensional lattice. This is however, easier said than done. Consequently a strictly empirical approach is undertaken, where by as many as possible parameters of crystal formation are searched (Carter *et al.*, 1988; Jancarik *et al.*, 1988), the most successful of which are then optimised to achieve the best crystals for diffraction analysis.

Crystals cannot grow in a saturated solution due to its state of equilibrium and therefore requires a non-equilibrium or super-saturated solution, which will provide the thermodynamic driving force for crystallisation.

In order to achieve crystal growth of a compound, a solution of the molecule must by some means be brought into the super-saturated state in which its subsequent return to equilibrium forces exclusion of solute molecules into the solid state, i.e. the crystal. The rate of crystal growth is some function of the distance of the solution from the equilibrium position at saturation. A crystal that forms far from equilibrium will grow very rapidly at first, and rapid growth is frequently associated with the occurrence of flaws and dislocations. Therefore, it is important to achieve the necessary degree of super-saturation for crystal growth but not too saturated that poor quality crystals are generated.

8.1.3 CRYSTALLISATION STRATEGY

The strategy of initiating crystallisation is to guide the system very slowly towards a state of reduced solubility by modifying the properties of the solvent. This is brought about by increasing the concentration of precipitating agents or by altering some physical property, such as temperature or pH. In this way, a limited degree of super-saturation and the promotion of specific bonding interactions between the molecules can be achieved. The precipitating agent may be a salt such as ammonium sulphate, an organic solvent such as ethanol or a highly soluble synthetic polymer such as poly (ethylene glycol). The three types of precipitants act to diminish the solubility of a protein by making its interactions with the solvent less favourable, mainly through competition with the protein for water. Protein molecules must bind to water in

order to remain soluble. However, when their interactions with water are lessened, they tend to associate with other protein molecules instead.

Vapour diffusion is the principle technique used in conjunction with the precipitants and physical parameters designed to propagate super-saturation of a protein solution via controlled evaporation or diffusion of the solvent (Hampel *et al.*, 1968) and was employed in this section of work. This approach allows the equilibrium of the protein/precipitant solution in a closed container with a larger aqueous reservoir, whose precipitant concentration is optimal for producing crystals; for example, the hanging drop method shown in Fig. 8.1.

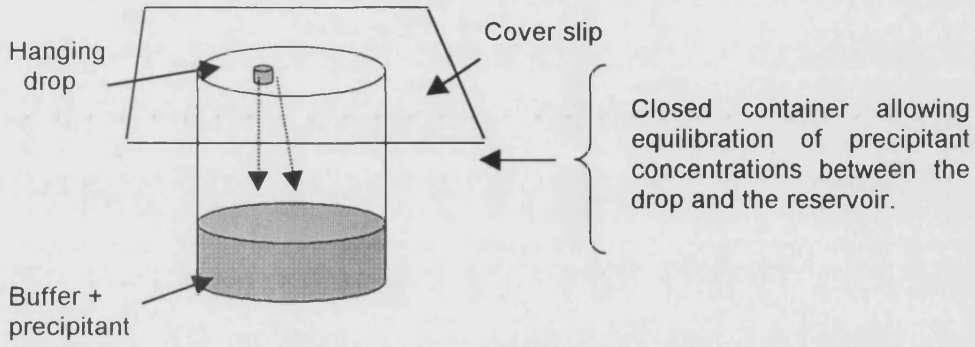


Fig. 8.1 Growing crystals by the hanging drop method.

The droplet hanging under the cover slip contains buffer, protein and precipitant at approx. 50% of the reservoir, thus allowing the net transfer of water from the protein solution to the reservoir until the precipitant concentration is the same in both solutions.

8.1.4 PROTEIN CRYSTALS

Macromolecular crystals are composed of approximately 50% solvent on average, though this may vary from 25 – 90% depending on the particular macromolecule (Matthews, 1968). The remaining volume is occupied by the protein or nucleic acid, so that the entire crystal is in many ways an ordered gel with extensive interstitial spaces through which solvent and other small molecules may freely diffuse. The hydrated protein surfaces that make up these interstitial spaces also act to create hydrogen bonds through which proteins can be held together. Therefore, to retain the integrity of the crystals, the crystals must not be allowed to dehydrate.

In entering the crystalline state from solution, individual molecules of the substance adopt one or only a few orientations. The resulting crystal is an orderly three-dimensional array of molecules. Each molecule or arrangement of molecules makes up a unit cell that is repeated throughout the crystal, which is described as the crystal lattice. The unit cell is an imaginary building block that merely provides a convenient mathematical device for describing the aggregation of molecules in a crystal. By convention, a unit cell is characterised by three vectors, a, b and c, and the interaxial angles between them, α, β, and γ.

8.1.5 COLLECTING X-RAY DATA

Fig. 8.2 depicts a simple portrayal of the collection of X-ray diffraction data. A crystal is mounted between an X-ray source and an X-ray detector so that it lies in the path of a narrow beam of X-rays.

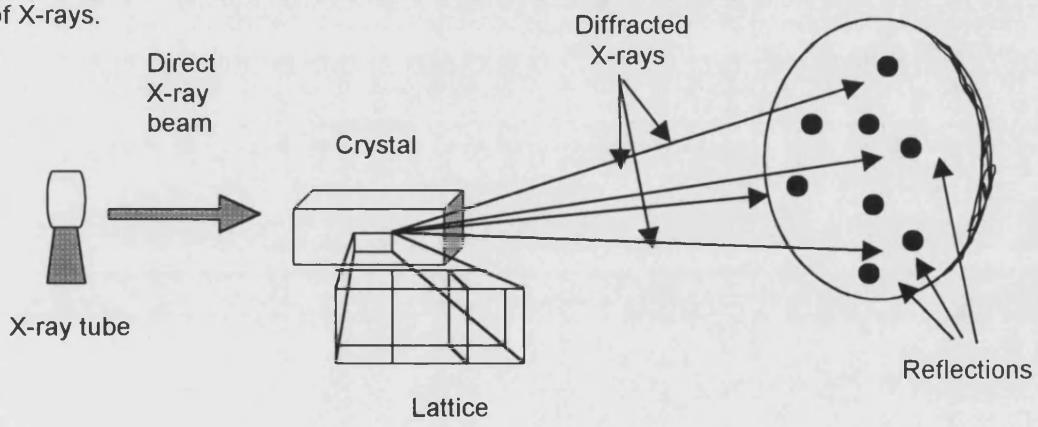


Fig. 8.2 Crystallographic data collection.

The crystal diffracts the source beam into many discrete beams, each of which produces a distinct spot (reflection) on the film. The positions and intensities of these reflections contain the information needed to determine molecular structures.

An optical scanner can then precisely measure the position and intensity position (which denotes the direction and strength of that particular beam respectively) of each reflection, and transmit it in digital form to a computer for analysis. These two parameters can then be entered into the computer programme to construct an image of the molecules in the unit cell.

8.1.5 MODEL BUILDING

Knowledge-based modelling of homologous proteins involves the superposition of multiple structures. However, in the case of a protein that has no crystallographic structural data but is thought to share a high structural homology with a protein of known structure, it is possible to obtain a three-dimensional image of the protein by basing it on the known structure. A computer-generated plot of the co-ordinated set, transformed to one with the best fit to the template structure, can be visually inspected. Such plots can be drawn as stereoscopic projections to aid visual perception of similarities and differences in the conformations of the molecules studied.

The aim of this section of work is primarily to deduce the structure of the GDH from *S. solfataricus* (although the author realises that this is a massive undertaking and would not be realised within the time constraints of this project). The chosen method to attaining this goal is single X-ray crystallography, and this therefore requires good quality protein crystals of GDH. Whilst trials were in progress, it was hoped that homology modelling would provide some insights into the actual structure, using the assumed high structural homology between the only known archaeal GDH structure (from *T. acidophilum*) (John *et al.*, 1994) and *S. solfataricus* GDH (Fig 4.21) despite their rather low sequence homology (Table Fig.4.2).

8.2 MATERIALS AND METHODS

8.2.1 MATERIALS

Structure screen kit, vacuum grease, sialysed glass cover slips, and Linbro cell tissue culture plates were all obtained from Molecular Dimensions Ltd., distributed by Stratech Scientific Ltd., Luton, Bedfordshire, U.K. Centricon Concentrators were provided by Millipore (UK) Ltd., Watford, Hertfordshire, U.K. Tris, ammonium sulfate, PEG 400 and PEG 8000 were obtained from Fisher Scientific U.K. Ltd., Loughborough, Leicestershire, U.K. Hepes and $MgCl_2$ were purchased from Sigma Chemical Company, Poole, Dorset, U.K.

8.2.2 EQUIPMENT

'In house' Siemens area detector and Microvax 4000, Silicon Graphics Indigo and an 'O'- a graphics package were used in conjunction with a SG Indigo for the visualisation of the electron density maps during the protein modelling work. The following software packages were also used: PROCHECK - a comprehensive package that checks the stereochemistry of the model, 3D-1D PROFILE – determines whether the model makes biochemical sense, GCG – University of Wisconsin Genetics Computer Group: a sequence analysis software package, and MOLSCRIPT – a programme to produce both detailed and schematic plots of proteins (Kraulis, 1991).

8.2.3 METHODS

8.2.3.1 PROTEIN PREPARATION

Protein was purified described in Chapter 5 with the addition of a final step using gel filtration chromatography with 50 mM Tris-HCl, pH 7, and 20 mM $MgCl_2$. It then was concentrated to 12.25 mg/ml using a centricon device, which involved the retention of protein via a membrane (nominal molecular weight limit, 50, 000 Daltons) on centrifuging the sample for approximately 20 min at 6000 x g.

8.2.3.2 SETTING UP CRYSTALLIZATION TRIALS

Firstly, the mouth of each well of a Linbro cell tissue culture plate was smeared with vacuum grease via a syringe. 1 ml of each trial buffer plus precipitant was placed in a well of the Linbro plate. On a pre-sialysed cover slip, 2 μl of protein sample at 12.25 mg/ml was then

mixed with 2 μl of the buffer plus precipitant from each well. The cover slip was then inverted and placed over its respective well and twisted slightly, in order to make an airtight seal with the aid of the vacuum grease. Once all the trials were set up and sealed, the whole plate was placed at a constant temperature of 22°C.

8.2.3.3 CRYSTAL ASSESSMENT

The scale of crystal quality was assessed using microscopic examination and a modified version of an arbitary scale developed by Carter *et al.*, (1988), such as that in Table 8.1, to grade each experiment.

RESULT	QUALITY
Clear drop	0
Precipitate	1
Unidentified something	2
Crystalline precipitate	3
Sea urchins, "starbursts"	4
Bundles of sticks	5
Needles	6
Plates/ chunky needles	7
Nice 3D crystals	8
3D crystals that diffract (not huge)	9
3D crystals, large, that diffract	10

Table 8.1 Scale of crystal quality

8.2.3.4 CRYSTAL MOUNTING

Crystals need to be mounted in such a way that they can be easily manipulated during intensity measurements. The method used here was to take up the crystal directly from the hanging drop into a thin-walled glass capillary tube. A drop or two of the mother liquor was also taken up on either side of the crystal so that an equilibrium atmosphere was maintained. Each end of the capillary tube was then sealed with beeswax and mounted on to the goniometer head. This is a device that holds the crystal in place on a camera and allows it to be oriented in the X-ray beam by means of translational and angular motions.

8.2.3.5 PROTEIN MODELLING

In order to begin the comparison of the structures of the thermophilic GDHs, the amino acid sequences were first aligned using the PILEUP programme in the GCG package. Information gained from this alignment concerning residue differences, gaps and insertions were then used to alter the GDH sequence of monomer A of *T. acidophilum* (Bright *et al.*, 1993) installed in the 'O' program (Jones *et al.*, 1991). Starting from the N-terminus of monomer A, residues were substituted for those indicated by the *S. solfataricus* GDH sequence (*mutate_replace*). Residues corresponding to insertions and deletions were then amended (*mutate_insert*; *mutate_delete*). These insertions and deletions were then accommodated into the *T. acidophilum* monomer structure by selecting the best piece of secondary structure from the 'O' database (*lego_loop*) to bridge the alteration and reassess the backbone conformation in that area. A rotamer of each inserted residue that optimally fitted the observed density was chosen from the 'O' database (*lego_side_chain*). The resulting monomer was assessed for the suitability of its main chain and side chain geometries using PROCHECK (Morris *et al.*, 1992) and the 3D-1D score (Luthy *et al.*, 1992).

8.3 RESULTS

8.3.1 CRYSTALLISATION TRIALS

8.3.1.1 Screening results

The purified protein was subjected to an initial random screen based on the INFAC (Carter *et al.*, 1988) and 'Magic 50' (Jancarik and Kim, 1991) approaches. These encompass all the factors deemed potentially important or proven to be successful in the crystallisation of a protein, such as a range of precipitants, cations, anions, cofactors and pH.

Once the initial condition(s) at which crystallisation of a protein has been achieved, the quality and size of the crystal should be improved (if needed). This is accomplished initially by optimising the primary crystallisation conditions, to produce larger crystals. Another option is crystal-seeding (Thaller *et al.*, 1981; Stura and Wilson, 1990) but this was not explored in this work.

The following screen results include the primary random screen (1) (Table 8.2), which produced several positive crystallisation conditions. The most successful trials were those of neutral to slightly alkaline pH with PEG 400 and 8000, and ammonium sulphate as the precipitating agents. The largest crystals produced were of the following respective dimensions: 1.0 x 0.1 x 0.1 mm and 0.3 x 0.1 x 0.1 mm. Therefore, the two further optimisation screens (2) and (3) (Tables 8.3 and 8.4) included variations of these conditions. However, in the third and final screen, the protein solution was diluted to approximately 6mg/ml, in order to try and slow down the rate of crystal growth. Some of these proved successful and also resulted in crystal formation, with the best-sized crystals being obtained from trial 3 and having the following dimensions, 1.0 x 0.2 x 0.2 mm, although they did not appear to be smooth and well formed. In all but a few successful cases (such as screen 3, trials 2 and 3). Crystals were obtained after 24 h; however, they were left for a period of a few days to ensure that they had reached maturity in terms of their size.

Photographic depictions of the crystal types obtained from these trials are shown in Figs 8.4a,b and c. These were obtained from the following trials: screen 1, trial 23; screen 1, trial 38; and screen 2, trial 14; respectively.

TRIAL	CONDITIONS			CRYSTAL QUALITY
	salt	buffer	precipitation agent	
1	0.02m Calcium chloride dihydratae	0.1M Na acetate trihydrate pH 4.6	30% (v/v) 2-methyl-2,4 pentanediol	0
2	0.2M Ammonium acetate	0.1M Na acetate trihydrate pH 4.6	30% (w/v) PEG 4000	1
3	0.2M Ammonium sulphate	0.1M Na acetate trihydrate pH 4.6	25% (w/v) PEG 4000	1/3
4	None	0.1M Na acetate trihydrate pH 4.6	2.0M Sodium formate	0
5	None	0.1M Na acetate trihydrate pH 4.6	2.0M Ammonium sulphate	2
6	None	0.1M Na acetate trihydrate pH 4.6	8% (w/v) PEG 4000	2
7	0.2M Ammonium acetate	0.1M tri-sodium citrate dihydrate pH 5.6	30% (w/v) PEG 4000	1/3
8	0.2M Ammonium acetate	0.1M tri-sodium citrate dihydrate pH 5.6	30% (v/v) 2-methyl-2,4-pentanediol	2
9	None	0.1M tri-sodium citrate dihydrate pH 5.6	20% (w/v) 2-propanol, 20% (w/v) PEG 4000	1
10	None	0.1M Na Citrate pH 5.6	1.0M Ammonium dihydrogen phosphate	2
11	0.2M Calcium chloride dihydrtae	0.1M Na acetate trihydrate pH 4.6	20% (v/v) 2-propanol	1
12	None	0.1M Na Cacodylate pH 6.5	1.4M Na acetate trihydrate	2
13	0.2M tri-sodium citrate dihyrate	0.1M Na Cacodylate pH 6.5	30% (v/v) 2-propanol	1
14	0.2M Ammonium sulphate	0.1M Na Cacodylate pH 6.5	30% (w/v) PEG 8000	1
15	0.2M Magnesium acetate tetrahydrate	0.1M Na Cacodylate pH 6.5	20% (w/v) PEG 8000	1/3
16	0.2M Magnesium acetate tetrahydrate	0.1M Na Cacodylate pH 6.5	30% (v/v) 2-methyl-2,4-pentanediol	0
17	None	0.1M Imidazole pH 6.5	1.0M Sodium acetate trihydrate	0
18	0.2M Sodium acatate trihydrate	0.1M Na Cacodylate pH 6.5	30% (w/v) PEG 8000	1
19	0.2M Zinc acatate dihydrate	0.1M Na Cacodylate pH 6.5	18% (w/v) PEG 8000	1/3
20	0.2M Cacium acaetate hydrate	0.1M Na Cacodylate pH 6.5	18% (w/v) PEG 8000	6
21	0.2M tri-sodium citrate dihydrate	0.1M Na Hepes pH 7.5	30% (v/v) 2-methyl-2,4-pentanediol	0
22	0.2M Magnesium chloride hexahydrate	0.1M Na Hepes pH 7.5	30% (v/v) 2-propanol	1
23	0.2M Calcium chloride dihydrate	0.1M Na Hepes pH 7.5	30% (v/v) 2-propanol	6/7
24	0.2M Magnesium chloride hexahydrate	0.1M Na Hepes pH 7.5	20% (v/v) PEG 400	6
25	0.2M tri-sodium citrate dihydrate	0.1M Na Hepes pH 7.5	20% (v/v) 2-propanol	1/3
26	None	0.1M Na Hepes pH 7.5	0.8M K, Na tartrate tetrahydrate	6
27	None	0.1M Na Hepes pH 7.5	1.5M Lithium sulphate monohydrate	6
28	None	0.1M Na Hepes pH 7.5	0.8M Na dihydrogen phosphate, 0.8M K dihydrogen phosphate monohydrate	0
29	None	0.1M Na Hepes pH 7.5	1.4M tri-sodium citrate dihydrate	3
30	None	0.1M Na Hepes pH 7.5	2% (v/v) PEG 400, 2.0M Amm sulphate	6/7
31	None	0.1M Na Hepes pH 7.5	10% (v/v) 2-proanol, 20% (w/v) PEG 400	1
32	None	0.1M Tris-HCl pH 8.5	2.0M Ammonium sulphate	5/6
33	0.2M Magnesium chloride hexahydrate	0.1M Tris-HCl pH 8.5	30% (w/v) PEG 4000	1
34	0.2M tri-sodium citrate dihydrate	0.1M Tris-HCl pH 8.5	30% (v/v) PEG 400	7
35	0.2M Lithium sulphate monohydrate	0.1M Tris-HCl pH 8.5	30% (w/v) PEG 4000	1
36	0.2M Ammonium acetate	0.1M Tris-HCl pH 8.5	30% (v/v) 2-propanol	1
37	0.2M Sodium acatate trihydrate	0.1M Tris-HCl pH 8.5	30% (w/v) PEG 4000	1
38	None	0.1M Tris-HCl pH 8.5	8% (w/v) PEG 8000	6
39	None	0.1M Tris-HCl pH 8.5	2.0M Ammonium dihydrogen phosphate	0
40	None	None	0.4M K, Na tartrate tetrahydrate	7
41	None	None	0.4M Ammonium dihydrogen phosphate	0
42	0.2M Ammonium sulphate	None	30% (w/v) PEG 8000	1
43	0.2M Ammonium sulphate	None	30% (w/v) PEG 4000	1
44	None	None	2.0M Ammonium sulphate	6/7
45	None	None	4.0M Sodium formate	6
46	0.05M Potassium dihydrogen phosphate	None	20% (w/v) PEG 8000	3
47	None	None	30% (w/v) PEG 1500	1
48	None	None	0.2M Magnesium formate	6/7
49	1.0M Lithium sulphate monhydrate	None	2% (w/v) PEG 8000	0/2
50	0.5M Lithium sulphate monohydrate	None	15% (w/v) PEG 8000	1

Table 8.2 Conditions and results of primary random screen (1)

BUFFERS	PRECIPITANTS AND CRYSTAL QUALITIES					
	2M $(NH_4)_2SO_4$	2M $(NH_4)_2SO_4$ 2% PEG 400	5% PEG 8000	10% PEG 8000	20% PEG 8000	30% PEG 8000
0.1M Hepes pH 7.0	1 0	2 0	3 0	4 6/7	5 3	6 3
0.1M Hepes pH 7	7 6	8 0	9 0	10 6	11 3	12 2
0.1M Tris-HCl pH 8.0	13 6/7	14 6/7	15 7	16 6	17 3	18 6
0.1M Tris-HCl pH 8.5	19 6/7	20 6	21 7	22 3	23 3	24 6/7

Table 8.3 Conditions and results of first optimisation screen (2)

The number of each trial is indicated in the top left-hand corner.

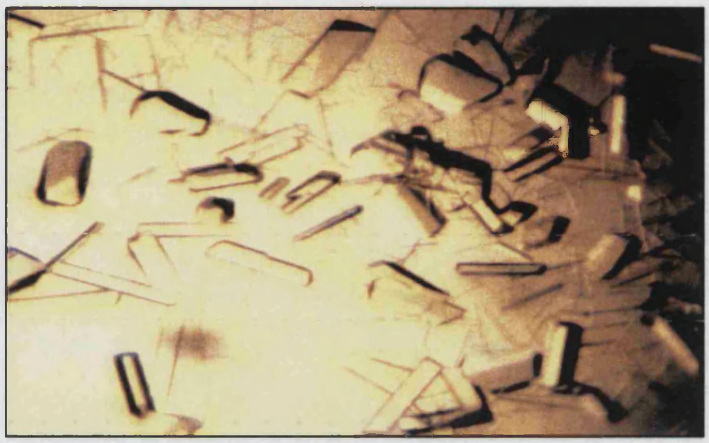
PRECIPITANTS	BUFFERS AND CRYSTAL QUALITIES					
	0.1M Tris-HCl pH 8.0	0.1M Tris-HCl pH 8.2	0.1M Tris-HCl pH 8.4	0.1M Tris-HCl pH 8.6	0.1M Tris-HCl pH 8.8	0.1M Glycine pH 9.0
2% PEG 8000	1 0	2 6	3 6/7	4 0	5 0	6 0
4% PEG 8000	7 0	8 7	9 0	10 0	11 0	12 0
6% PEG 8000	13 6/7	14 7	15 6/7	16 0	17 6	18 2/3
8% PEG 8000	19 6/7	20 7/6	21 6/7	22 6	23 6	24 3

Table 8.4 Conditions and results of second optimisation screen (3)

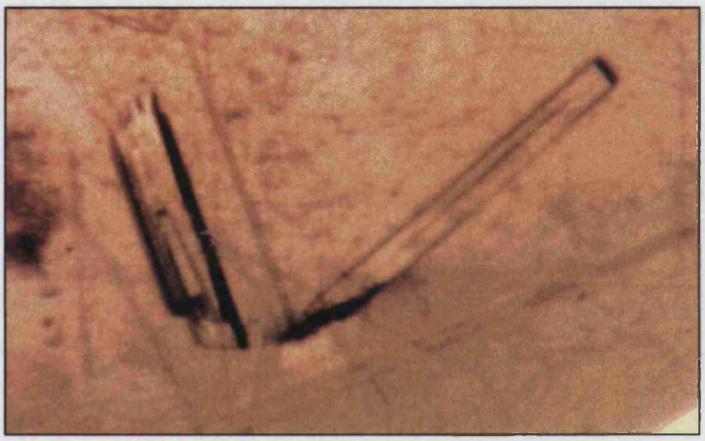
The number of each trial is indicated in the top left-hand corner.

The crystals shown in Fig. 8.3a were used in X-ray crystallography trials using the in-house rotating anode X-ray source. They were analysed with the crystal detector set at 30 cm. Only 13 contiguous frames of diffraction data were collected, oscillating the crystal 1.0° each frame. Diffraction densities were measured to a resolution of 7.0 Å. However, the data obtained were not of sufficient resolution in order to be able to obtain the space group and other detailed structural information, although it was possible to determine the unit cell dimensions and the actual crystal symmetry using these data (8.3.1.2).

(a)



(b)



(c)

Fig. 8.3 Crystals of *S. solfataricus* GDH.

(a) from screen 1, trial 23, (b) from screen 1, trial 38 and (c) from screen 2, trial 14.

8.3.1.2 The crystal density and unit cell contents

Analysis of a wide range of protein crystals (Matthews, 1968) has determined the fraction of the crystal volume that is occupied by solvent to be about 50 %, but values ranging from 30 – 70 % have also been found. The sufficient restriction imposed by this range enables the direct determination of the probable number of molecules in the crystallographic asymmetric unit from the molecular mass of the protein, the space group and the unit cell dimensions of the crystal. The results may indicate the presence of solvent of crystallisation in the crystal or it may also indicate that the crystal cannot possibly accommodate what it was expected to. Therefore, crystal density measurements are essential in the early stages of a crystal structure analysis.

The Matthews coefficient (V_m) (Matthews, 1968) is used to estimate solvent content:

$$V_m = \frac{\text{Cell volume (Å}^3\text{)}}{\text{Mass of monomer (Daltons) x n (no. of molecules in the asymmetric unit)}}$$

The volume of the unit cell in Å^3 is deduced from the diffraction data and its dimensions were determined to be 68, 92 and 138 Å (≈ 863328 Å^3), as mentioned earlier. The relative molecular mass of GDH was calculated from the sequence of each monomer and found to be 40849 Daltons, and *n* is the number of asymmetric units per unit cell to be determined. Solvent content is then calculate by:

$$V_{solv} \approx 1 - \frac{1.23}{V_m}$$

However, as the number of molecules in the asymmetric unit is not known for sure, it is then estimated by knowing that the Matthews coefficient is usually between 1.66 and 4, and solvent content is therefore usually close to 50% (i.e. between 30 and 70 %). Therefore:

Number of tetramers in unit cell	V_m Equation	V_m (Å/Dalton)	Corresponding solvent content (%)
1	$\frac{863328}{4 \times 40849}$	5.28	77
2	$\frac{863328}{8 \times 40849}$	2.64	53
3	$\frac{863328}{12 \times 40849}$	1.76	30

Therefore, the assumption of the presence of 50 % solvent in the crystal, provides two tetramers (molecules) in the unit cell.

Also, observation of the unit cell dimensions (68, 92 and 138 Å) and knowledge of the interaxial angles, $\alpha = \beta = \gamma = 90°$, will provide information on the crystal class, which in this case is orthorhombic. Orthorhombic symmetry dictates the presence of four asymmetric units per unit cell. Therefore, each asymmetric unit must contain a dimer in order to correlate with the unit cell data, which contains two tetramers. However, in order for this to be feasible, a non-crystallographic 2-fold axis of symmetry within the dimers which meet to form the native tetramer on a crystallographic 2-fold axis is required, as in *T. Acidophilum* GDH (John *et al.*, 1994).

8.3.2 PROTEIN MODELLING

8.3.2.1 Model

The initial alignment of the thermophilic GDH sequences shown in Fig. 8.4, was the basis for the final model depicted in Fig. 8.8.

```
   1                                                    50
S.s GDH  ...MKAIIVK PPNAGVQVKD VDEKKLDSYG KIKIRTIYNG ICGTDREIVN
T.a GDH  MTEQKAIVTD APKGGVKYTT IDMPEPEHYD .AKLSPVYIG ICGTDRGEVA

    51                                                   100
S.s GDH GKLTLSTLPK GKDFLVLGHE AIGVVEESYH G..FSQGDLV MPVNRRGCGI
T.a GDH GALSFTYNPE GENFLVLGHE ALLRVDDARD NGYIKKGDLV VPLVRRP.GK

   101                                                   150
S.s GDH CRNCLVGRPD FCETGEFG.. EAGIHKMDGF MREWWYDDPK YLGKIPKS.I
T.a GDH CINCRIGRQD NCSIGDPDKH EAGITGLHGF MRDVIYDDIE YLVKVEDPEL

   151                                                   200
S.s GDH EDIGILAQPL ADIEKSIEEI LEVQKRVPVW TCDDGTLNCR KVLVVGTGPI
T.a GDH GRIAVLTEPL KNVMKAF.EV FDVVSKRSIF FGDDSTLIGK RMVIIGSGSE

   201                                                   250
S.s GDH GVLFTLLFRT YGLEVWMANR REPTEVEQTV IEETKTNYYN SSNGYDKLKD
T.a GDH AFLYSFAGVD RGFDVTMVNR HDETENKLKI MDEFGVKFAN Y......LKD

   251                                                   300
S.s GDH SVGKFDVIID ATGADVNILG NVIPLLGRNG VLGLFGFSTS GSV.PLDYKT
T.a GDH MPEKIDLLVD TSGDPTTTF. KFLRKVNNNG VVILFGTNGK APGYPVDGED

   301                                                   350
S.s GDH LQEIVHTNKT IIGLVNGQKP HFQQAVVHLA SWKTLYPKAA KMLITKTVSI
T.a GDH IDYIVERNIT IAGSVDAAKI HYVQALQSLS NWNRRHPDAM KSIITYERSR

                351                         376
S.s GDH NDEKELLKVL REKEHGEIKI RILWE*
T.a GDH PK.....PTY SSRNHTER*. ......
```

Fig. 8.4 **Sequence alignment of the GDHs from *S. solfataricus* (S.s) and *T. acidophilum* (T.a). Compiled using the PILEUP programme in the GCG package.**

The sequences appear to align very well, apart from a noticeable gap present in the *T. acidophilum* sequence at about residue number 250 and the insertions at the C-terminus of the *S. solfataricus* enzyme. This alignment was used to model the structure of *S. solfataricus* GDH as described in section 8.2.3.5.

8.3.2.2 Assessment of model quality

The model quality was assessed using a 3D-1D score (Luthy *et al.*, 1992), which gives a measure of the biochemical sense of the structure, and a PROCHECK analysis (Morris *et al.*, 1992) in which a Ramachandran plot displays residues that exhibit acceptable stereochemistry and highlights the outlying residues.

The 3D-1D score was correlated for the monomer (Fig. 8.5) and gave very low scores i.e. those that are approaching or below zero, for the region 320 – 335aa, which is the extended C-terminal portion present in the *S. solfataricus* GDH sequence (Fig. 8.4). Several of the residues present in this loop are also highlighted in the Ramachandran plot (Fig. 8.6). Thus showing that they are also not in positions to achieve the most favourable protein conformation, which possesses the lowest free energy. Subsequent manipulation of the C-terminal loop region via *lego_loop* and *lego_side_chain* operations and re-submission of the co-ordinates into the 3D-1D profile program, gave much more favourable scores for the C-terminal region (Fig. 8.7)

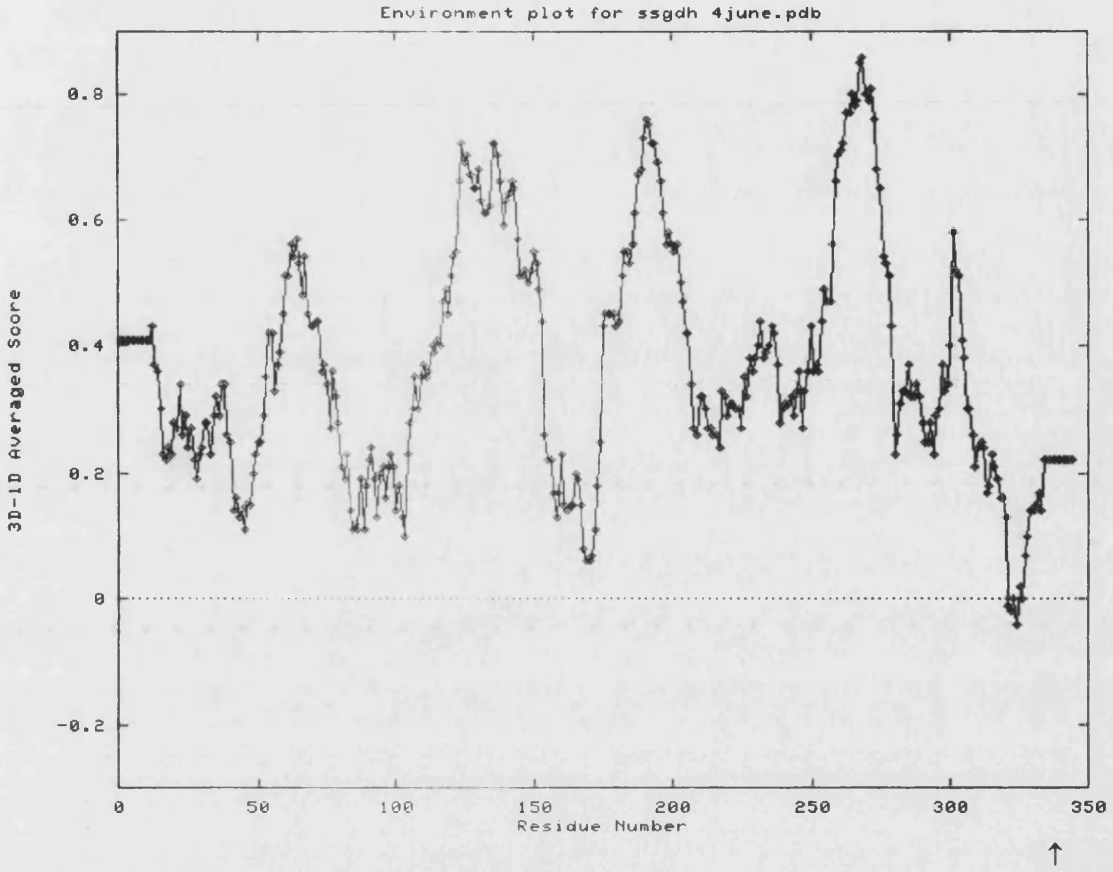
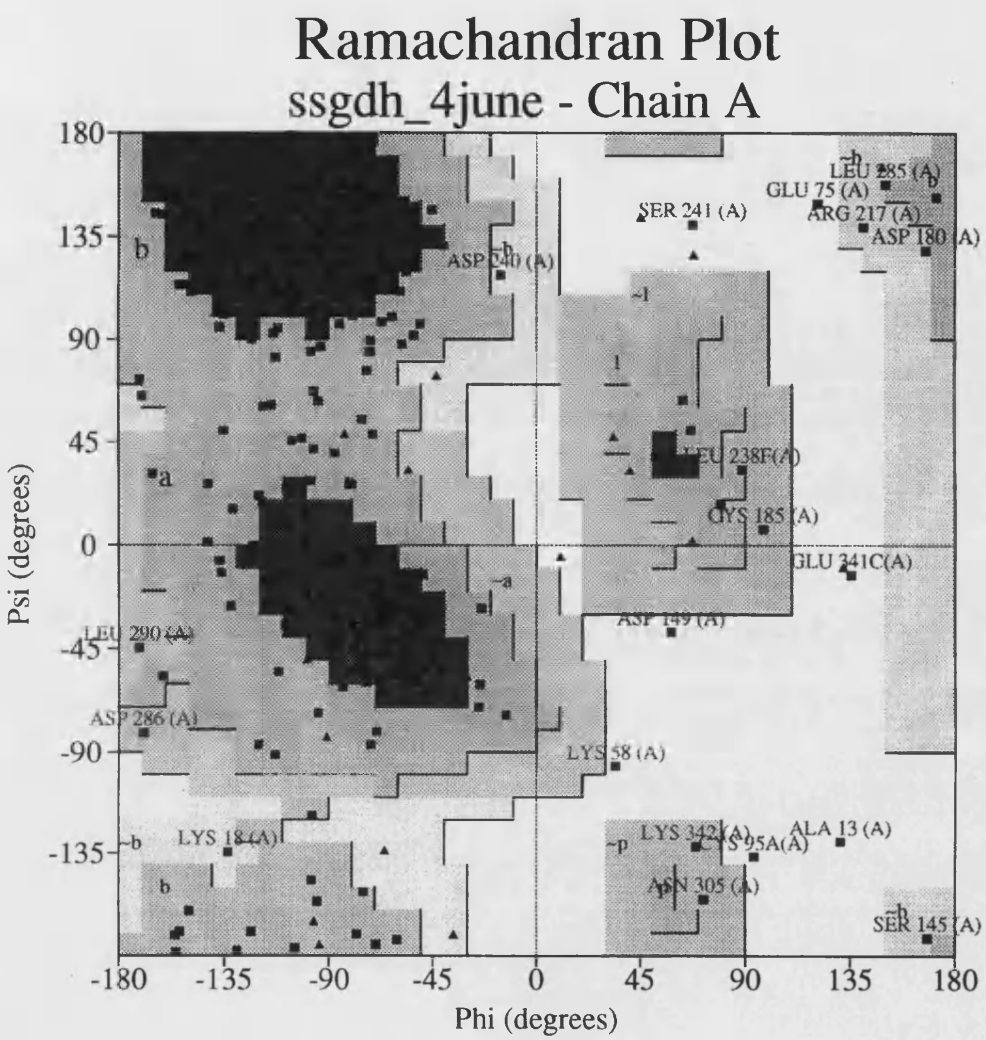


Fig. 8.5 The 3D-1D score for monomer A.

Region of low score indicated by arrow.



Plot statistics

Residues in most favoured regions [A,B,L]	214	70%
Residues in additional allowed regions [a,b,l,p]	75	24%
Residues in generously allowed regions [~a,~b,~l,~p]	12	4%
Residues in disallowed regions	7	2%
Number of non-glycine and non-proline residues	308	100%
Nunber of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	36	
Number of proline residues	15	
Total number of residues	361	

Fig. 8.6 Ramachandran plot for monomer A.

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

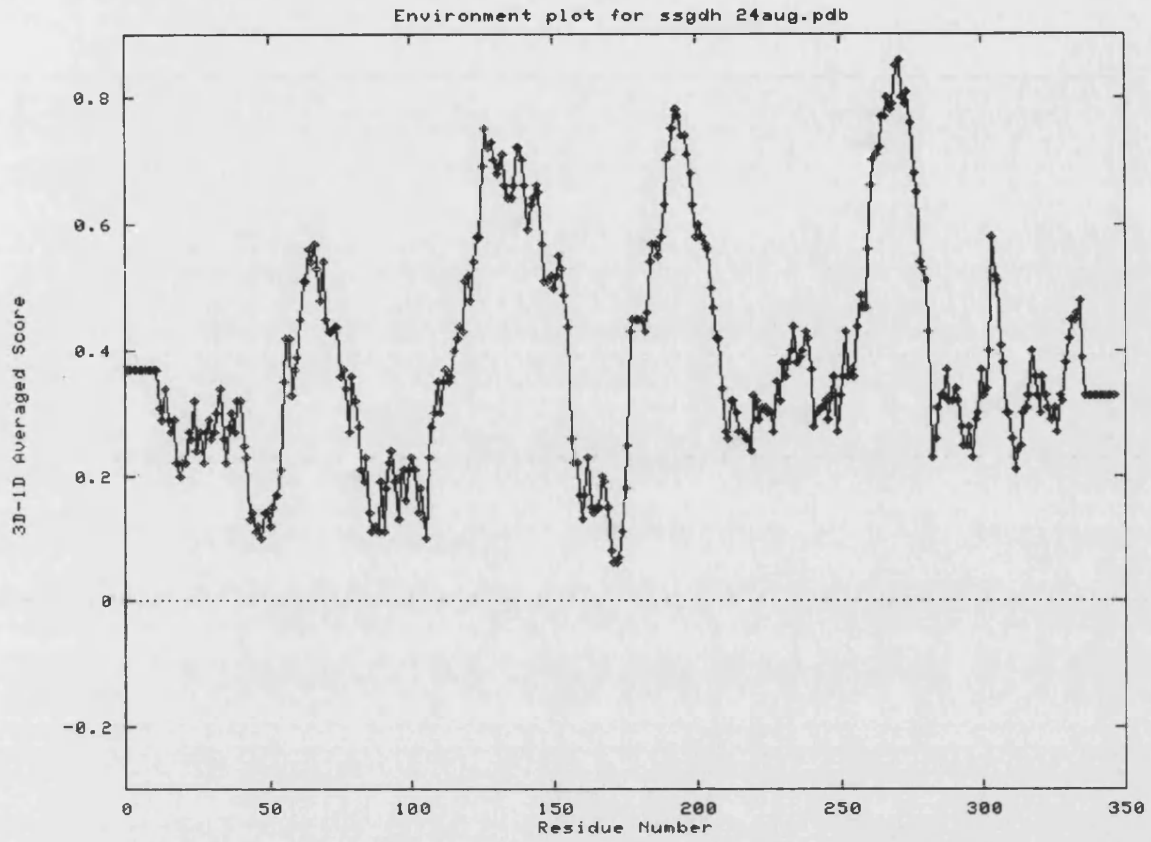


Fig. 8.7 Second 3D-1D score for the refined monomer A.

Following the creation, assessment and enhancement of the *S. solfataricus* GDH model it is necessary to view it and compare it to the *T. acidophilum* GDH template, in order to observe any structural changes, which may elucidate its higher thermostability. The structures of both these GDH monomers are shown superimposed in Fig. 8.8, as a stereoview. Large proportions of the structures are superimposable, except for two regions. The first is at the C-terminal loop that extends out from the catalytic domain, which is longer in *S. Solfataricus* GDH than the equivalent loop observed in *T. acidophilum* GDH and extends further out into solution. The second significant difference is found at the structural zinc-binding lobe, in which the lobe in *S. solfataricus* GDH has been somewhat flattened in respect to the hook-like structure of the *T. acidophilum* GDH structural lobe.

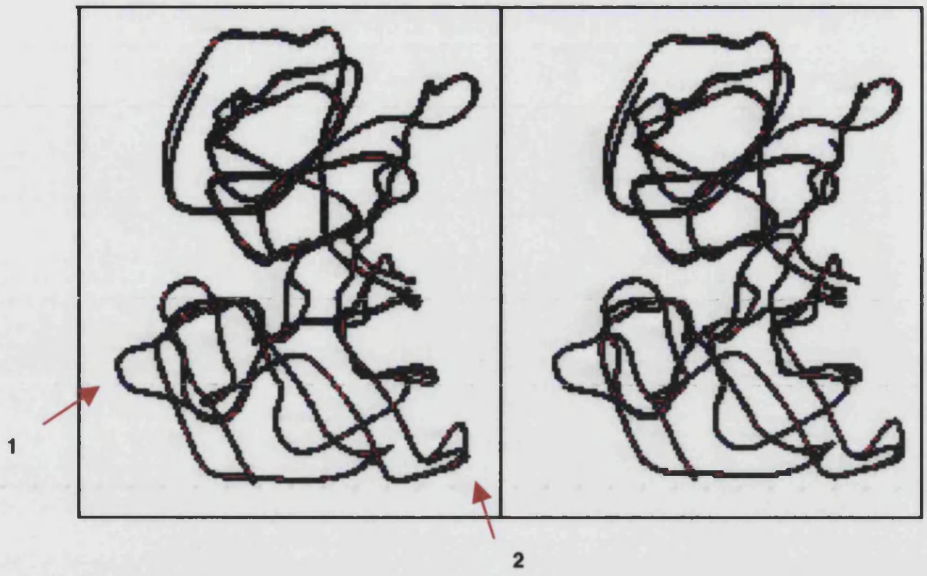


Fig. 8.8 **Stereoview of the superimposed Cα tracings of *S. solfataricus* GDH monomer A (blue) and the monomer A of *T. acidophilum* GDH (red), drawn using MOLSCRIPT.**

Arrows indicate the major deviations in both structures, namely the C-terminal (1) and the loop in the catalytic domain structural zinc binding lobe (2).

Despite the minor differences in structure outlined in Fig. 8.8, the differences in secondary structure are required to be visualised in more detail in order to observe whether the shorter loops associated with withstanding higher physiological temperatures are seen in *S. solfataricus* GDH when compared to those found in *T. acidophilum* enzyme. Therefore, the alignment displayed in Fig. 8.9 was compiled. Fig. 8.9 shows that there is no significant reduction in loop regions (those regions denoted by S and T) and as expected due to the nature of the production of the model, all the helical and sheet regions are highly conserved among the two GDH structures.

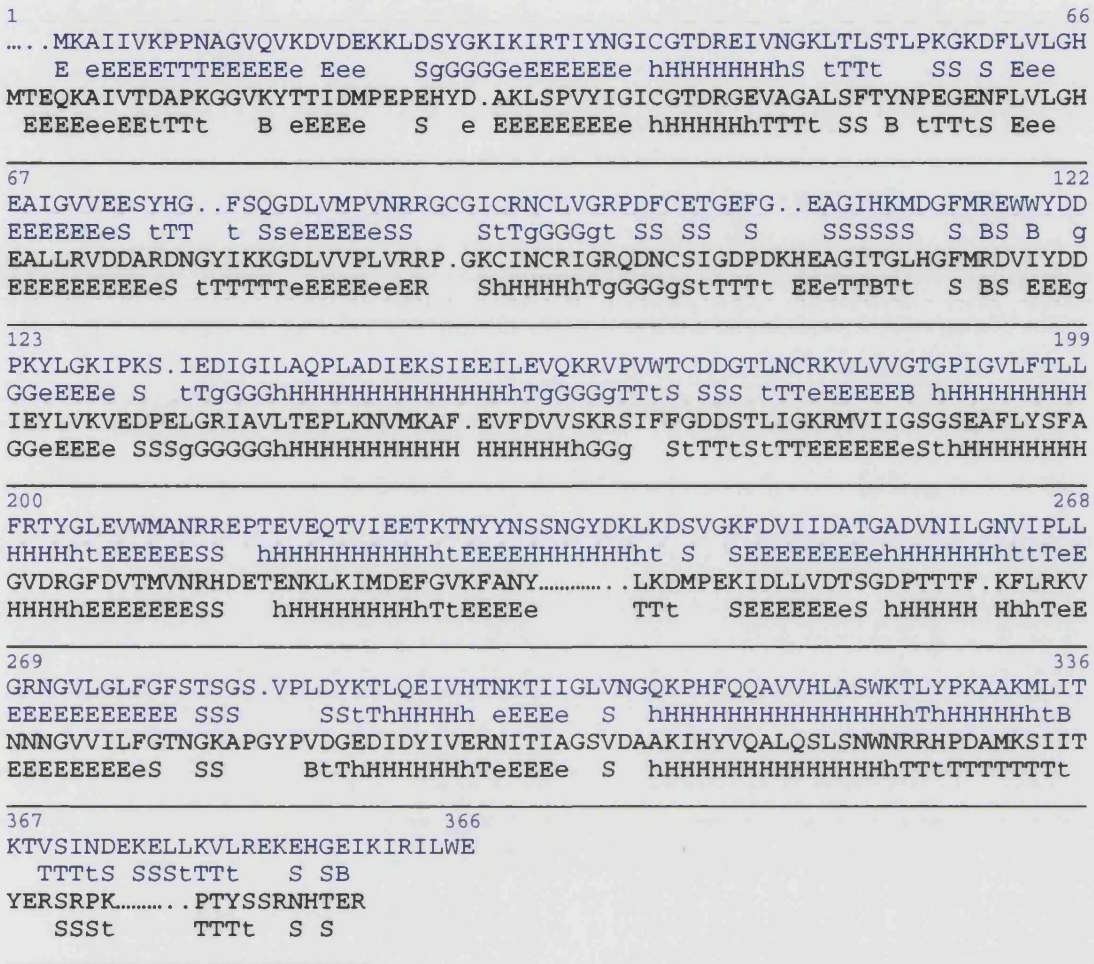


Fig. 8.9 **Secondary structural alignment of *S. solfataricus* (in blue) and *T. acidophilum* GDH (black).**

Secondary structure (extended Kabsch/Sander) key codes = B - residue in isolated beta-bridge, E - extended strand, participates in beta-ladder, G - 3-helix (3/10 helix), H - 4-helix (alpha-helix), I - 5-helix (pi-helix), S – bend, T - hydrogen-bonded turn, e - extension of beta- strand, g - extension of 3/10 helix and h - extension of alpha-helix. Numbers indicate the position of residues in the *S. solfataricus* GDH sequence.

8.4 DISCUSSION

8.4.1 CRYSTALLIZATION TRIALS

Although the trials carried out in this work were successful in terms of the production of crystals, the rapid growth of these crystals may have caused the occurrence of flaws and dislocations, which have led to the failure to obtain high-resolution X-ray diffraction data. This could be combated by slowing down the growth process through the reduction of the protein concentration (as attempted in screen 3) and/or perhaps the reduction of the growth temperature, although, this was found to have detrimental effects on the crystal formation of *T. acidophilum* GDH (Rossjohn, 1990). Another reason for poor crystal quality is the presence of microheterogeneity within the protein sample, as encountered with the trials carried out with the *T. acidophilum* GDH (Mackness, 1991). However, this was later found to be the result of the presence of a multitude of impurities (Rossjohn, 1994). The apparent efficiency of the purification procedure (Chapter 5) and the inclusion of a further chromatography step should have prevented this from occurring. However, it is still possible for microheterogeneity to result from other sources, for instance, the presence, absence, or variation of a bound group or metal ion, the variation in the aggregation or oligomer state of the protein to association/dissociation and conformational instability due to the dynamic nature of the molecule (McPherson, 1990).

Other possibilities to improve crystal quality, despite the potential creation of microheterogeneities, is the inclusion of substrates, coenzymes, inhibitors and metal ions. The presence of substrate, coenzyme or inhibitor will serve to fix an enzyme in a more stable and compact form and may lead to greater structural homogeneity due to the reduction of lability. The inclusion of the metal ion may aid in maintaining certain structural features of the molecule in view of the requirement of divalent cations for activity (chapter 3).

Despite the limited success of the crystallisation trials carried out in this work, these results are being used as the basis for exploring more favourable crystallisation conditions (by H. Walden and Prof. G. Taylor at St. Andrews), in order to obtain the crystal structure of *S. solfataricus* GDH.

8.4.2 PROTEIN MODELLING

The assumption of high structural homology of the two GDH enzymes, based on the data from secondary prediction analyses and sequence alignments, is one of the limitations to this exercise. However, the similarities in substrate specificity (chapter 6), dual cofactor specificity (with preferences for $NADP^+$), zinc co-ordinations, conserved sequence sites and also perhaps their shared ancient origin (which would denote their phylogenetic proximities and therefore their possible emergence from a common ancestor, which bears the same catalytic traits) make this more reasonable.

The expected correlations between local hydrophobicity, length of surface loops, degree of packing and concealment of thermolabile amino acids, and the increasing thermostability, of the *T. acidophilum* GDH and *S. solfataricus* GDH, cannot be fully realised from a model. This is due to the degree of subjectivity applied when producing such a model, especially in the absence of any high-resolution crystallographic data for the *S. solfataricus* GDH. This is clearly seen in the manual alteration of the α-carbon backbone and side chain orientations of the *S. solfataricus* monomer A (Fig. 8.4) in the region 320 – 335 (which corresponds to the C-terminal loop in the catalytic domain), in order to give a more favourable conformation and 3D-1D score. However, a low 3D-1D score was also observed in this region for a model of *T. acidophilum* GDH (based on the LADH structure and crystallographic data) (Rossjohn, 1994) and this was improved when re-correlated with the complete tetramer, thus suggesting the participation of this region in subunit/subunit interactions. However, this region does lie close to the loop extending from the nucleotide-binding domain. Therefore, perhaps oligomer formation leads to a conformational change in the monomers, causing a reduction in steric hindrance between the domains by these two loop structures, or indeed creates a stable interaction between these two loops. Consequently, manipulation of the C-terminal region of the *S. solfataricus* monomer A may have abolished its role in interacting with the nucleotide-binding domain.

The Ramachandran plot for monomer A revealed that the majority of the residues are in the most favoured regions and the additionally allowed regions (93.9 %). The conformation of Glu75 was consistent with the stereochemistry adopted by Asp75 in the *T. acidophilum* GDH,

which is also deemed to be unusual. Asp75 in the *T. acidophilum* GDH was thought to be part of a surface loop that was involved in crystal packing and therefore may provide an explanation of its strained conformation (Rossjohn, 1994).

The structural changes outlined by the superimposed models (Fig. 8.8) show the difference in the C-terminal loop conformation of *S. solfataricus* GDH. Its attainment of a significantly more favourable 3D-1D score, suggests that it was involved in steric hindrance, presumably by its proximity to the loop extending from the nucleotide-binding domain. Introduction of the co-ordinates for the complete tetramer of *S. solfataricus* GDH into the 3D-1D profile may produce different 3D-1D scores. However, without knowing the actual C-terminal conformation, we can only speculate on its involvement.

Also highlighted by the superimposition of the GDH models is the difference in structural zinc-binding lobes. The more hook-like structure of the lobe observed in the *T. acidophilum* GDH may help to maintain its structural integrity by "cradling" the zinc atom, which is thought to provide the protein with its degree of thermostability (Rossjohn,1994). However, in view of the higher thermostability of the *S. solfataricus* enzyme, instead of this structure being even more pronounced, the loop is considerably flatter. The hook conformation may simply be due to the absence of a fourth cysteine residue in *T. acidophilum* GDH usually involved in co-ordinately binding a zinc atom (John *et al.*, 1994). This has been replaced by a proline residue (Pro95) and therefore caused *T. acidophilum* GDH to employ an aspartate residue (Asp115) as the fourth ligand.

However, the most likely role for these regions could be in maintaining the quaternary structure through subunit-subunit interactions, as hypothesised for the *T. acidophilum* GDH. This hypothesis was based on the role of the equivalent regions of inter-subunit contacts in LADH (John *et al.*, 1994), where the interconnecting helix (residues 154 – 174) and the C-terminal loop of *T. acidophilum* GDH, were both found to be longer than the equivalent regions in LADH (residues 175 – 184). This enabled the loop to interact with structural lobes from two of the other monomers and provided an additional site for subunit-subunit association that was not observed in the LADH dimer. Therefore, further extension of the C-terminal loop seen in *S.*

solfataricus would presumably strengthen this interaction and participate to provide its greater observed thermostability.

The 3D-1D scores and Ramachandran plot (Figs 8.5 and 8.6) of the resulting model of *S. solfataricus* GDH show that all the residues are in favourable environments. It is of course homologous to the *T. acidophilum* GDH structure. This would therefore also influence the appearance of the structural sequence alignment of the two archaeal GDHs, in which no discernible differences in loop lengths were observed (Fig. 8.9) and consequently does not appear to fit with one of the proposed structural determinants of protein thermostability. However, as stated by Jaenicke, (1996), it is expected that the phenomenon of thermal stability can be achieved by more than one general mechanism and in fact, the two major structural discrepancies that were found in *S. solfataricus* GDH, at the C-terminal loop and the structural zinc-binding lobe may be found to play a role in increased thermostability.

9 Chapter Nine

General discussion

9.1 SUMMARY

In summary, this body of work has detailed the isolation of a hyperthermophilic enzyme from its host. This has also lead to the cloning and sequencing of the gene that codes for this enzyme in order to see if it groups with other known sequences in the gene databases. Expression of the gene product in a mesophilic organism then amplified its levels for subsequent characterisation studies and application in a biotechnological system.

Sequencing the *S. solfataricus* GDH revealed it to belong to a sub-group of the pyridine-nucleotide-dependent dehydrogenase superfamily, that is the medium to long-chain-zinc-requiring alcohol/polyols/dehydrogenases/reductases (MDR). This includes a number of medium-chain alcohol, polyol, threonine, archaeal glucose dehydrogenase and sorbitol dehydrogenases (and glutathione-dependent formaldehyde dehydrogenase)as well as quinone oxidoreductase, enoyl reductases and erythronolide synthases (Jörnvall *et al.*, 1987; Persson *et al.*, 1994; Edwards *et al.*, 1996). This also highlighted regions of high sequence conservation, namely the nucleotide-, structural zinc-, and the catalytic zinc-binding sites, that are also present in the other fully characterised archaeal GDH from *T. acidophilum*.

The recombinant enzyme was expressed in *E. coli*, achieving 7/8-fold the expression levels found in the host organism, athough, this could be vastly improved if carried out in a *rec* A positive strain. It was purified with the ease of a heat step and affinity chromatography to yield a specific activity of 50 U/mg. This was an improvement on the purification of the native enzyme, which attained this value using three chromatography steps. The purified protein was found to be a homomeric tetramer of 164 kDa.

This GDH was found to require divalent cations for activity and was able to use both NAD^+ and $NADP^+$ to achieve differing rates of aldose oxidation. GDH has a broad substrate specificity but was found to require its substrates to be aldopyranoses with equatorially oriented non-H substituents at the C-2, C-3 and C-4 positions. It achieved the highest rates of oxidation

with D-Glucose, giving the following kinetic data: K_M with NAD^+ 1500 μM, with $NADP^+$ 1300 μM; V_{MAX} with NAD^+ 110 μmoles/min/mg, with $NADP^+$ 70 μmoles/min/mg. The K_M data of the cofactors with D-Glucose as a co-substrate were 900 and 50 μM for NAD^+ and $NADP^+$ respectively.

This GDH was also found to possess greater thermostability than the GDH from *T. acidophilum,* although it also displayed some degree of instability when incubated at elevated temperatures in the presence of substrates. GDH was also found to display a temperature optimum of 90°C, although, this value was affected by the disclosure of additional studies revealing the possible occurrence of reversible conformational changes in GDH at temperatures exceeding 75°C. The activity of the GDH was also found to be enhanced at high pH. This is presumed to be due to the positive ionisation effects on the residues and zinc ion involved in binding and catalysis, as well as the favourable shift in the equilibrium brought about by this change in hydroxide levels.

The application of the *S. solfataricus* GDH in conjunction with a hyperthermophilic hydrogenase from *P. furiosus* in the *in vitro* enzymatic hydrogen production system established by Woodward *et al.*, (1996) proved to have an overall advantage compared with the system involving the *T. acidophilum* GDH. It was able to achieve greater rates and yields of hydrogen production at higher temperatures and lower enzyme concentrations using the *S. solfataricus* enzyme. However, it was also subject to its own limitations with respect to the substrate and product concentrations, which brought about inhibition of GDH activity and stability. Additional work on the variables of the system as a whole revealed other limitations, specifically pH and temperature. These factors have been found to have a profound affect on the stability of the cofactor used, $NADP^+$ (Wong and Whitesides, 1981; Walsh *et al.*, 1983; Robb *et al.*, 1992; Daniel and Danson, 1994), and for this reason were therefore thought to be responsible for the lower yields and rates.

Preliminary crystallisation trials with purified GDH successfully led to crystal growth using a random screening kit, in which success was achieved at higher pH values (all above its isoelectric point of pH 7.0) and in the presence of the precipitating agents PEG and ammonium sulphate. The resulting crystals were however fairly small (1.0 x 0.1 x 0.1 mm) and not of suitable quality in order to obtain high-resolution x-ray diffraction data, although, it was possible

to use them in diffraction experiments to obtain estimations of the unit cell, which showed that the crystal belonged to the orthorhombic space group with unit cell dimensions a = 68 Å, b = 92 Å, c = 138 Å and $\alpha = \beta = \gamma = 90°$, indicating that there were two tetramers of GDH in the asymmetric unit.

Attempts to obtain a map image of the *S. solfataricus* GDH structure in the absence of any crystallographic data was fairly limited in its usefulness, due to the subjective nature of the homology modelling procedure. However, the exercise proved some success in that superimposing the *S. solfataricus* GDH amino acid sequence on to the crystal structure of a *T. acidophilum* GDH monomer resulted in a highly homologous model, with the exception of the C-terminal region. This region had to be manually manipulated in the monomer-state, in order to produce a structure that was feasible in terms of its main- and side-chain geometries. However, previous modelling studies on the monomer structure of *T. acidophilum* GDH (Rossjohn, 1994), also revealed unfavourable geometries in its C-terminus. Yet these were corrected when the structure was determined crystallographically in a tetrameric state, thus suggesting a possible role for the C-terminus in subunit binding. Therefore, perhaps the extended C-terminus seen in *S. solfataricus* GDH provides a greater interaction with its fellow subunits, in view of its higher temperature habitat than that of *T. acidophilum* GDH.

9.2 FUTURE WORK

In order to improve or expand on the data obtained during this study, several lines of further study can be employed, if further time was available. Production of more recombinant enzyme would in turn facilitate the completion of the kinetic data for GDH in order to obtain V_{MAX} values for the cofactors (NAD^+ and $NADP^+$) with D-Glucose as a co-substrate. Subsequent use of this data to calculate the catalytic efficiencies and the assumption of the common ancestor of the archaebacteria to be most closely related to the earliest form of life (Woese *et al.*, 1990), may lead to further information on the evolution of dual cofactor specificity and preferential usage of NAD/P(H) when compared to other pyridine-nucleotide-dependent dehydrogenases across all three domains of life.

If more time was available more information on the GDH structure and mechanism of catalysis and substrate binding could be obtained. For example:

- GDH could be subjected to proton microprobe analysis (bombarding a GDH crystal with a high energy proton beam) and flame atomic absorption spectrometry analysis (aspiration of a GDH protein sample into the flame-nebulizer system of a spectrometer). This would confirm whether, the putative catalytic and structural zinc-binding sites do indeed involve the co-ordination of two zinc atoms per GDH monomer.
- Indeed GDH protein samples could be subjected to thermostability studies in the presence of increasing amounts of the chelating agent, EDTA. In the *T. acidophilum* GDH this brought about the selective removal of the putative structural zinc at low temperatures, which is more solvent accessible (Rossjohn, 1994). This would also deduce whether the putative structural zinc was responsible for its level of thermostability.
- Steady state tryptophan fluorescence studies in the presence of a range of $NADP^+$ concentrations and more in depth product inhibition studies (using gluconolactone as opposed to gluconic acid) would provide information on whether GDH exhibits a compulsory order mechanism and binds $NADP^+$ first.
- Completetion of structure determination by X-ray crystallography perhaps through the inclusion of substrates in the crystallisation trials.

References

A

Aguilar, C.F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., & Pearl, L.H. (1997) *J. Mol. Biol.* **271**, 789 – 802. Crystal Structure of the β-Glycosidase from the Hyperthermophilic Archaeon *Sulfolobus solfataricus* Resilience as a Key Factor in Thermostability.

Ahern, T.J. & Kilbanov, A.M. (1985) *Science* **228**, 1280 – 1284. The Mechanism of Irreversible Enzyme Inactivation at 100°C.

Akanuma, S., Yamagishi, A., Tanaka, N., & Oshima, T. (1998) *Protein Sci.* **7**, 698 - 705. Serial increase in the thermal stability of 3-isopropylmalate dehydrogenase from *Bacillus subtilis* by experimental evolution.

Al-Karadaghi, S., Cedergren-Zeppenzauer, E.S., & Hövmoller, S. (1994) *Acta. Cryst.* **D50**, 793 - 807. Refined crystal structure of liver alcohol dehydrogenase-NADH complex at 1.8 Å resolution.

Ammendola, S., Raia, C.A. Caruso, C., Carmardella, L., D'Auria, S., De Rosa, M., & Rossi, M. (1992) *Biochemistry* **31**, 12514 - 12523. Thermostable NAD^+-Dependent Alcohol Dehydrogenase from *Sulfolobus solfataricus*: Gene and Protein Sequence Determination and Relationship to Other Alcohol Dehydrogenases.

Anfinsen, C.B. (1973) *Science* **181**, 223 - 229. Principles that govern the folding of protein chains.

Aravind, L., Tatsuov, R.L., Wolf, Y.I., Walker, R., & Koonin, E. (1998) *Trends Genetic.* **14**, 11, 442 - 444. Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles.

Argos, P., Rossmann, M., Grau, U., Zuber, H., Frank, G., & Tratschin, J.D. (1979) *Biochemistry* **8**, 5698 - 5703. Thermal stability and protein structure.

Arnold, F.H. (1998) *Acc. Chem. Res.* **31**, 125 - 131. Design by directed evolution.

Arnold, F.H. & Volkov, A.A. (1999) *Curr. Opin. Chem. Biol.* **3**, 1, 54 – 59. Directed evolution of biocatalysts.

Ashokkumar, M. (1998) *Int. J. Hydrogen Energy* **23**, 6, 427 - 438. An overview on semiconductor particulate systems for photoproduction of hydrogen.

Auer, J., Lechner, K., & Böck, A. (1989) *Can. J. Microbiol.* **35**, 200 - 204. Gene organisation and structure of two transcriptional units from Methanococcus coding for ribosomal proteins and elongation factors.

Avigad, G., Alroy, Y., & Englard, S. (1968) *J. Biol. Chem.* **243**, 8, 1936 - 1941. Purification and Properties of a Nicotinamide Adenine Dinucleotide Phosphate-linked Aldohexose Dehydrogenase from *Gluconobacter cerinus.*

B

Baker, P.J., Britton, K.L., Rice, D.W., Rob, A., & Stillman, T.J. (1992a) *J. Mol. Biol.* **228**, 662 - 671. Structural consequences of sequence patterns in the fingerprint region of the nucleotide binding fold. Implications for nucleotide specificity.

Baker, P.J., Britton, Engel, P.C., Farrants, G.W., Lilley, K.S., Rice, D.W., & Stillman, T.J. (1992b) *Proteins* **12**, 75 - 86. Subunit assembly and active site location in the structure of glutamate dehydrogenase.

Barns, S.M., Delwiche, C.F., Palmer, J.D., & Pace, N.R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 9188 - 9193. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences.

Bergquist, C. & Parkin, G. (1998) *Inorg. Chem.* **38**, 422 - 423. Modelling the catalytic cycle of Liver alcohol dehydrogenase: synthesis and structural characterisation of a four-co-ordinate zinc ethoxide complex and determination of relative Zn-OR *versus* Zn-OH bond energies.

Biellman, J-F., Lapinte, C., Haid, E., & Wiemann, G. (1979) *Biochemistry* **18**, 7, 1212 – 1216. Structure of lactate dehydrogenase inhibitor generated from coenzyme.

Bischoff, K.M. & Rodwell, V.W. (1996) *J. Bacteriol.* **178**, 19 - 23. A reductase from *Haloferax volcanii*: purification, characterisation and expression in *Escherichia coli.*

Blecher, O., Goldman, S., & Mevarech, M (1993) *Eur. J. Biochem.***216**, 199 - 203. High expression in *Escherichia coli* of the gene coding for dihydrofolate reductase of the extremely halophilic archaebacterium *Haloferax volcanii.* Reconstitution of the active enzyme and mutation studies.

Blochl, E., Rachel, R., Buggraf, S., Hafenbradl, D., Jannasch, H.W., & Stetter, K.O. (1997) *Extremophiles* **1**, 14 - 21. *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper limit for life to 113°C.

Böhm, G. & Jaenicke, R. (1994) *Prot. Eng.* **7**, 213 - 220. A structure-based model for the halophilic adaptation of dihydrofolate reductase from *Halobacterium volcanii.*

Bonete, M-J., Pire, C., Llorca, F., & Camacho, M.L. (1996) *FEBS Lett.* **383**, 227 - 229. Glucose dehydrogenase from the halophilic Archaeon *Haloferax mediterranei*: enzyme purification, characterisation and N-terminal sequence.

Bradford, M.M. (1976) *Anal. Biochem.* **72**, 248 – 254. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilising the Principle of Protein-Dye Binding.

Bragger, J.M., Daniel, R.M., Coolbear, T., & Morgan, H.W. (1989) *Appl. Microbiol. Biotechnol.* **31**, 556 - 561. Very stable enzymes from extremely thermophilic archaebacteria and eubacteria.

Bränden, C-I. (1977) *Pyridine nucleotide-dependent dehydrogenases*, 325 - 338, Ed. Horst Sund, Walter de Gruyter, Berlin, New York. Functional significance of the structure of liver alcohol dehydrogenase.

Bright, J.R. (1991) *Ph. D. Thesis.* University of Bath.

Bright, J.R., Byrom, D., Danson, M.J., Hough, D.W., & Towner, P. (1993) *Eur. J. Biochem.* **211**, 549 - 554. Cloning, sequencing and expression of the gene encoding glucose dehydrogenase from the thermophilic Archaeon *Thermoplasma acidophilum.*

Brock, T.D., Brock, K.M., Belly, R.T., & Weiss, R.L. (1972) *Arch. Microbiol.* **84**, 54 – 68. *Sulfolobus* – A New Genus of Sulfur-Oxidizing Bacteria Living at Low pH and High Temperature.

Brown, J.R. & Doolittle, W.F. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 2441 - 2445. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications.

Brown, S.H., Sjøholm, C., & Kelly, R.M. (1993) *Biotech. Bioeng.* **41**, 878 - 886. Purification and Characterisation of a Highly Thermostable Glucose Isomerase Produced by the Extremely Thermophilic Eubacterium, *Thermotoga maritima.*

Bryant, F. O. & Adams, M.W.W. (1989) *J. Biol. Chem.* **264**, 9, 5070 - -5079. Characterization of Hydrogenase from the hyperthermophilic Archaebacterium, *Pyrococcus furiosus.*

Budgen, N. & Danson, M.J. (1986) *FEBS* **196**, 2, 207 – 210. Metabolism of glucose via a modified Entner-Doudoroff pathway in the thermoacidophilic archaebacterium *Thermoplasma acidophilum.*

Bult, C.J., White, O., Olsen, G.J., Zhou, L.X., Fleischman, R,D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., *et al.* (1996) *Science* **273**, 1058 - 1073. Complete genome sequence of the methogenic archaeon *Methanococcus jannaschii.*

Burdette, D.S., Vielle, C., & Zeikus, J.S. (1996) *Biochem. J.* **316**, 115 - 122. Cloning and expression of the gene encoding the Thermoanaerobacter ethanolicus 39E secondary-alcohol dehydrogenase and biochemical characterization of the enzyme.

C

Campbell, D.P., Carper, W.R., & Thompson, R.E. (1982) *Arch. Biochem. Biophys.* **215**, 1, 289 - 301. Bovine liver glucose dehydrogenase: isolation and characterization.

Carter, Jr., C.W., Baldwin, E.T., & Frick, L. (1988) *J. Cryst. Growth* **90**, 60 - 73. Statistical design of experiments for protein crystal growth and the use of a precrystallization assay.

Castresana, J. & Saraste, M. (1995) *TIBS* **20**, 443 - 447. Evolution of energetic metabolism: the respiration-early hypothesis.

Cedergren-Zeppenzauer, E. (1983) *Biochemistry* **22**, 5761 - 5772. Crystal-determination of reduced nicotinamide adenine dinucleotide complex with horse liver alcohol dehydrogenase maintained in its apo conformation by zinc-bound imidazole.

Cendrin, F., Chroboczek, J., Zaccai, G., Eisenburg, H., & Mevarech, M. (1993) *Biochemistry* **32**, 4308 - 4313. Cloning, sequencing, and expression in *Escherichia coli* of the gene coding for malate dehydrogenase of the extremely halophilic archaebacterium *Haloarcula marismortui.*

Chan, M.K., Mukund, S., Kletzin, A. Adams, M.W.W., & Rees, D.C. (1995) *Science* **267**, 1463 – 1469. Structure of a Hyperthermophilic Tungstopterin Enzyme, Aldehyde Ferredoxin Oxidoreductase.

Cheetam, P.S.J. (1998) *J. Biotechnol.* **66**, 3 - 10. What makes a good biocatalyst?

Chenault, H.K. & Whitesides, G.M. (1987) *Appl. Biochem. Biotech.* **14**, 2, 147 – 197. Regeneration of nicotinamide cofactors for use in organic synthesis.

Chien, A., Edgar, D.B., & Trela, J.M. (1976) *J. Bacteriol.* **127**, 1550 - 1557. Deoxyribonucleic acid polymerase from the extremophile *Thermus aquaticus.*

Ciulla, R.A., Burggraf, S., Stetter, K.O., & Roberts, M.F. (1994) *Appl. Env. Microbiol.* **60**, 10, 3660 - 3664. Occurrence and role of Di-*myo*-iositol-1,1'-phosphate in *Methanococcuss igneus.*

Cleton-Jansen, A-M., Goosen, N., Wenzel, T.J., & Van De Putte, P. (1988) *J. Bacteriol.* **170**, 2121 - 2125. Cloning of the gene encoding quinoprotein glucose dehydrogenase from *Acinetobacter calcoaceticus* : evidence for the presence of a second enzyme.

Cleton-Jansen, A-M., Goosen, N., Vink, K., & Van De Putte, P. (1989) *Mol. Gen. Genet.* **217**, 430 - 436. Cloning and characterization and DNA sequencing of the gene encoding the Mr 50,000 quinoprotein glucose dehydrogenase from *Acinetobacter calcoaceticus.*

Cline, A.L. & Hu, S.L. (1965) *J. Biol. Chem.* **240**, 11, 4488 – 4492. The Isolation of Three Sugar Dehydrogenases from a *Pseudomonad.*

Colombo, S., D'Auria, S., Fusi, P., Zecca, L., Raia, C.A., & Tortora, P. (1992) *Eur. J. Biochem.* **206**, 349 - 357. Purification and characterization of a thermostable carboxypeptidase from the extreme thermophilic archaebacterium *Sulfolobus solfataricus.*

Connaris, H., Chaudhuri, J.B., Danson, M.J., & Hough, D.W. (1999) *Biotechnol Bioeng.* **64**, 1, 38-45. Expression, reactivation and purification of enzymes from *Haloferax volcanii* in *Escherichia coli.*

Connaris, H., West, S.M., Hough, D.W., & Danson, M.J. (1998) *Extremophiles* **2**, 61 - 66.Cloning and overexpression in *Escherichia coli* of the gene encoding citrate synthase from the hyperthermophilic Archaeon *Sulfolobus solfataricus.*

Consalvi, V., Chiraluce, R., Politi, L., Pasquo, A., DeRosa, M., & Scandurra, R. (1993) *Biochim. Biophys. Acta* **1202**, 207 - 215. Glutamate dehydrogenase from the thermoacidophilic archaebacterium *Sulfolobus solfataricus* studies on thermal and guanidine-dependent inactivation.

Copeland, H.F. (1938) *Quart. Rev. Biol.* **13**, 383.

Cowan, D.A. (1992) *Trends Biotechnol.* **10**, 315 - 323. Biotechnology of the Archaea.

Crameri, A., Raillard, S.A., Bermudez, E. & Stemmer, W.P. (1998) *Nature* **391**, 288 - 291. DNA shuffling of a family of genes from diverse species accelerates directed evolution.

Creighton, T.E. (1990) *Biochem. J.* **270**, 1 -16. Protein folding.

D

Daly, M.J., Ouyang, L., Fuchs, P., & Minton, K.W. (1994) *J. Bacteriol.* **176**, 12, 3508 - 3517. *In Vivo* Damage and *recA*-Dependent Repair of Plasmid and Chromosomal DNA in the Radiation-Resistant Bacterium *Deinicoccus radiodurans.*

Daniel, R.M. & Danson, M.J. (1995) *J. Mol. Evol.* **40**, 559 - 563. Did primitive micro-organisms use nonhem iron proteins in place of NAD/P?

Danson, M.J. (1989) *Can. J. Microbiol.* **35**, 58 - 64. Central metabolism of the archaebacteria: an overview.

Danson, M.J. & Hough, D.W., (1992) *Biochem. Soc. Symp.* **58**, 7 - 21. The enzymology of archaebacterial pathways of central metabolism.

Danson, M.J. & Hough, D.W., (1998) *Trends Microbiol.* **6**, 8, 307 – 314. Structure, function and stability of enzymes from the Archaea.

Danson, M.J., Hough, D.W., Russel, R.J.M., Taylor, G.L., & Pearl, L. (1996) *Prot. Eng.* **9**, 8, 629 – 630. Enzyme thermostability and thermoactivity.

Dawson, R.M.C., Elliot, D.C., Elliot, W.H., & Jones, K.M. eds. (1986) *Data for Biochemical Research*, 3rd edn., Oxford Science Publications.

Delboni, L.F., Mande, S.C., Rentire-Delrue, F., Mainfroid, V., Turley, S., Vellieux, F.M.D., Martial, J.A., & Hol, W.G.J. (1995) *Protein Sci.* **4**, 2594 - 2604. Crystal structure of recombinant triosephosphate isomerase from *Bacillus stearothermophilus.* An analysis of potential thermostability factors in sx isomerases with known three dimensional structures points to the importance of hydrophobic interactions.

De Bont, J.A.M. (1998) *Trends Biotechnol.* **16**, 493 – 499. Solvent-tolerant bacteria in biocatalysis.

De Long, E.F. (1998) *Curr, Opin, Genet. Dev.* **8**, 649 - 654. Everything in moderation: Archaea as "non-extremophiles".

De Long, E.F. (1998) *Science* **280**, 542 - 543. Archaeal means and extremes.

De Rosa, M., Gambacorta, A., Millonig, G., & Bu'Lock, J.D. (1974) *Experientia* **30**, 866 - 868. Convergent characters of extremely thermophilic acidophilic bacteria.

De Rosa, M., Gambacorta, A., & Bu'Lock, J.D. (1975) *J. Gen. Microbiol.* **86**, 156 – 164. Extremely Thermophilic Acidophilic Bacteria Convergent with *Sulfolobus acidocaldarius.*

De Rosa, M., Gambacorta, A., Nicolaus, B., Giardina, P., Poerio, E., & Buonocore, V. (1984) *Biochem. J.* **224**, 407 – 414. Glucose metabolism in the extreme thermoacidophilic archaebacterium *Sulfolobus solfataricus.*

Dill, K.A. (1990) *Biochem.* **29**, 31, 7133 - 7155. Dominant forces in protein folding.

Doolittle, W.F. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 8797 - 8799. At the core of the Archaea.

Doolittle, W.F. (1999) *Science* **284**, 2124 - 2128. Phylogenetic classification and the universal tree.

Doolittle, W.F. & Brown, J.R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 6721 - 6728. Tempo, mode the progenote, and the universal root.

Dong, G., Vielle, C., Savchenko, A., & Zeikus, J.G (1997) *Appl. Environ. Microbiol.* **63**, 9, 3569 - 3576. Cloning, Sequencing, and Expression of the Gene Encoding Extracellular α-Amylase from *Pyrococcus furiosus* and Biochemical Characterization of the Recombinant Enzyme.

Dröge, M., Pühler, A., & Selbitschka, W. (1998) *J. Biotech.* **64**, 75 - 90. Horizontal gene transfer as a biosafety issue: A natural phenomenon of public concern.

Dym, O., Mevarech, M., & Sussman, J.L. (1995) *Science* **267**, 1344 – 1346. Structural Features That Stabilize Halophilic Malate Dehydrogenase from an Archaebacterium.

E

Edgell, D.R. & Doolittle, W.F. (1997) *BioEssays* **19**, 1, 1 - 4. Archaebacterial genomics: the complete genome sequence of *Methanococcus jannaschi.*

Edwards, K.J., Barton, J.D., Rossjohn, J., Thorn, J.M., Taylor, G.L., & Ollis, D.L. (1996) *Arch. Biochem. Biophys.* **328**, 1, 173 - 183. Structural and sequence comparisons of Quinone oxidoreductase, ζ-Crystallin, and Glucose and Alcohol dehydrogenases.

Egerer, P., Günter, H., & Simon, H. (1982) *Biochim. Biophys Acta* **703**, 149 - 157. On the hydrogen-deuterium exchange reaction catalyzed by the soluble hydrogenase from *Alcaligenes eutrophus* H16 in the free and immobilized state.

Eisenthal, R. & Cornish-Bowden, A. (1974) *Biochem. J.* **139**, 3, 715 - 720. The direct linear plot. A new graphical procedure for estimating enzyme kinetic parameters.

Eklund, H., Bränden, C-I., & Jörnvall, H. (1976) *J. Mol. Biol.* **102**, 61 – 73. Structural Comparisons of Mammalian, Yeast and Bacillar Alcohol Dehydrogenases.

Eklund, H., Horjales, E., Jörnvall, H. Bränden, C-I., & Jeffery, J. (1985) *Biochemistry* **24**, 8005 - 8012. Molecular aspects of functional differences between alcohol and sorbitol dehydrogenases.

Eklund, H., Nordström, B., Zeppenezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B-O., Tapia, O., & Bränden, C-I., (1976) *J. Mol. Biol.* **102**, 27 - 59. Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution.

Eklund, H., Samama, J-P., & Jones, A. (1984) *Biochemistry* **23**, 5982 - 5996. Crystallographic investigations of nicotinamide adenine dinucleotide binding to horse liver alcohol dehydrogenase.

Ensely, B.D., Gibson, D.T., & Laborde, A.L. (1982) *J. Bacteriol.* **149**, 948 – 954.

Ensely, B.D., Ratzkin, B.J., Osslund, T.D., Simon, M.J., Wackett, L.P. & Gibson, D.T. (1983) *Science* **222**, 167 – 169. Expression of Naphthalene Oxidation Genes in *Escherichia coli* Results in the Biosynthesis of Indigo.

Entner, N.& Doudoroff, M. (1951) *J. Biol. Chem.* **196**, 853 – 863. Glucose and gluconic acid oxidation of *Pseudomonas saccharophila.*

Everse, J., Zoll, E.L., Kahan, L., & Kaplan, N.O. (1971) *Bioorg. Chem.* **1**, 207.

F

Fawcett, C.P., Ciotti, M.M., & Kaplan, N.O. (1961) *Biochim. Biophys. Acta* **54**, 210 - 212. Inhibition of dehydrogenase reactions by a substance formed from reduced diphosphopyridine nucleotide.

Feinstein, S.I., Chernajovsky, Y., Chen, L., Maroteaux, & Mory, Y. (1983) *Nuc. Acids Res.* **11**, 9, 2927 - 2941. Expression of human interferon genes using the *recA* promoter of *Escherchia coli.*

Fiala, G. & Stetter, K.O. (1986) *Arch. Microbiol.* **145**, 56 - 61. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C.

Fife, T.H. & Przystas, T.J. (1986) *J. Amer. Chem. Soc.* **108**, 4631 - 4636. Divalent metal ion catalysis in amide hydrolysis. The hydrolysis of N-Acylimidazoles.

Fitzgibbon, S., Choi, A.J., Miller, J.H., Stetter, K.O., Simon, M.I., Swanson, R., & Kim, U.J. (1997) *Extremophiles* **1**, 36 - 51. A fosmid-based genomic map and identification of 474 genes of the hyperthermophilic archaeon *Pyrobaculum aerophilum.*

Frey, B. & Suppman, B. (1995) *Biochemica* **2**, 34 - 35. Demonstration of the Expand PCR System's greater fidelity and higher yields with a lacI-based fidelity assay.

Fujinaga, M., Berthet-Colominas, C., Yaremchuk, A.D., Tukalo, M.A., & Cusack, S. (1993) *J. Mol. Biol.* **234**, 222 - 233. Refined crystal structure of the seryl-tRNA synthetase from *Thermus thermophilus* at 2.5 Å resolution.

G

Gambacorta, A., Trincone, A., Nicolaus, B., Lama, L., & De Rosa, M. (1994) *System. Appl. Microbiol.* **16**, 4, 518 - 527. Unique features of lipids of Archaea.

Garg, N., Galaev, I.Y., & Mattiasson, B. (1996) *J. Mol. Recognition* **9**, 259 - 274. Dye-Affinity Techniques for Bioprocessing: Recent Developments.

Gerday, C., Aittlaeb, M., Arpigny, J.L., Baise, E., Chessa, J.P., Garsoux, G., Petrescu, I., Feller, G. (1997) *Biochim. Biophys. Acta* **1342**, 119 – 131. Psychrophilic enzymes: a thermodynamic challenge.

Gerike, U., Danson, M.J., Russell, N.J., & Hough, D.W. (1997) *Eur. J. Biochem.* **248**, 49 -57. Sequencing and expression of the gene encoding a cold-active citrate synthase from an Antartic bacterium, strain DS2-3R.

Giardina, P., De Biasi, M-G., De Rosa, M., Gambacorta, A., & Buonocore, V. (1986) *Biochem. J.* **239**, 517 - 522. Glucose dehydrogenase from the thermoacidophilic archaebacterium *Sulfolobus solfataricus.*

Gogarten, J.P.. Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J., Manolson, M.F., Poole, R.J., Date, T., Oshima, T., Konishi, J., Denda, K., & Yoshida, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6661 - 6665. Evolution of the vacuolar H^+-ATPase: implications for the origin of eukaryotes.

Gottesman, S. (1981) *Cell* **23**, 1 -2. Genetic Control of the SOS System in *E. coli.*

Greenbaum, E. (1984) *Photobiochem. Photophys.* **8**, 323 - 332. Biophotolysis of water: the light saturation curves.

Groves, J.T. & Olsen, J.R. (1985) *Inorg.Chem.* **24**, 18, 2715 - 2717. Models of zinc-containing proteases - rapid amide hydrolysis by an unusually acidic Zn^{2+}-OH_2 complex. TI:

Gudas, L.J. & Mount, D.W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 12, 5280 - 5284. Identification of the *recA* (tif) gene product of *Escherichia coli.*

Gudas, L.J. & Pardee, A.B. (1975) *Proc. Nat. Acad. Sci. USA* **72**, 6, 2330 - 2334. Model for Regulation of *Escherichia coli* DNA Repair Functions.

Gudas, L.J. & Pardee, A.B. (1976) *J. Mol. Biol.* **101**, 459 - 477. DNA Synthesis Inhibition and the Induction of Protein X in *Escherichia coli.*

Gupta, M.N. (1992) *Eur. J. Biochem.* **203**, 25 – 32. Enzyme function in organic solvents.

Gupta, R.S. & Golding, G.B. (1996) *Trends Biochem. Sci.* **21**, 166 - 173. The origin of the eukaryotic cell.

H

Hallewell, R.A. & Emtage, S. (1980) *Gene* **9**, 27 - 47. Plasmid vectors containing the tryptophan operon promoter suitable for efficient regulated expression of foreign genes.

Hampel, A., Labanauskas, M., Conners, P.G., Kirkegard, L., RajBhandary, U.L., Sigler, P.B., & Brock, R.M. (1968) *Science* **162**, 1384 - 1387. Single crystals of transfer RNA from formylmethionine and phenylalanine transfer RNA's.

Hanukoglu, I. & Gutfinger, T. (1989) *Eur. J. Biochem.* **180**, 479 - 484. cDNA sequence of adrenodoxin reductase. Identification of NADP-binding sites in oxidoreductases.

Harrison, D.C. (1931) *Biochem. J.* **25**, 1016 - 1027. CXIII. Glucose Dehydrogenase: A New Oxidising Enzyme From Animal Tissues.

Harrison, D.C. (1932) *Biochem. J.* **25**, 1295 - 1299. CLVI. The Product of the Oxidation of Glucose by Glucose Dehydrogenase:

Heilman, H.J., Mågert, H.J., & Gassen, H.G. (1988) *Eur. J. Biochem.* **174**, 485 - 490. Identification and isolation of glucose dehydrogenase genes of *Bacillus megaterium* M1286 and their expression in *Escherichia coli.*

Hensel, R. (1993) *New Compr. Biochem.* **26**, 209 - 221.The Biochemistry of archaea (archaebacteria editors, M. Kates, D.J. Kushner, A.T. Matheson.

Hensel, R. & König, H. (1988) *FEMS Microbiol. Lett.* **49**, 75 - 79. Thermoadaptation of methogenic bacteria by intracellular ion concentration.

Herbert, R.A. (1992) *Trends Biotechnol.* **10**, 395 – 401. A perspective on the biotechnological potential of extremophiles.

Hochstein, L.I. (1988) *in Halophilic Bacteria* **2**, 67 - 83. Rodriguez-Valera, F., ed.), CRC Press Inc., Boca Raton, U.S.A.

Horii, T., Ogawa, T., & Ogawa, H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1, 313 - 317. Organization of the *recA* gene of *Escherichia coli*

Horikoshi, K. (1996) *FEMS Microbiol. Revs.* **18**, 259 – 270. Alkaliphiles – from an industrial point of view.

Hough, D.W. & Danson, M.J. (1999) *Curr. Opin. Chem. Biol.* **3**, 1, 39 – 46. Extremozymes.

Hugenholtz, P. & Pace, N.R. (1996) *Trends Biotechnol.* **14**, 190 - 196. Identifying microbial diversity in the natural environment: a molecular phylogenetic approach.

Humayun, M.Z. (1998) *Molecular Microbiology* **30**, 5, 905 - 910. SOS and Mayday: multiple inducible mutagenic pathways in *Escherichia coli.*

I

Ingledew, J.W. (1982) *Biochim. Biophys. Acta* **693**, 2, 89 - 117. Thiobacillus ferrooxidans. The bioenergetics of an acidophilic chemolithotroph.

Inoue, T., Kumar, S.N., Kamachi, T., & Okura, I. (1999) *Chem. Lett.* **2**, 147 - 148. Hydrogen evolution from glucose with the combination of Glucose dehydrogenase and Hydrogenase from *A. eutrophus* H16.

Iyer, V.N. & Szybalski, W. (1964) *Science* **145**, 55 - 58. Mitomycins and Porfiromycin: Chemical Mechanism of Activation and Cross-linking of DNA.

Iwabe, N., Kuma, K-I., Hasegawa, M., Osawa, S., & Miyata, T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9355 - 9359. Evolutionary relationships of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes.

J

Jaenicke, R. (1991) *Eur. J. Biochem.* **202**, 715 - 728. Protein stability and molecular adaptation to extreme conditions.

Jaenicke, R. & Zavodsky, P. (1990) *FEBS Lett.* **268**, 344 - 349. Proteins under extreme physical conditions.

Jancarik, J. & Kim, S-H. (1991) *J.Appl. Cryst.* **24**, 409 - 411.. Sparse matrix sampling: a screening method for crystallization of proteins

Jany, K-D., Ulmer, W., Fröschle, M., & Pfleiderer, G. (1984) *FEBS* **165**, 1, 6 - 10. Complete amino acid sequence of glucose dehydrogenase from *Bacillus megaterium.*

John, J., Crennel, S.J., Hough, D.W., Danson, M.J., Taylor, G.L. (1994) *Structure* **2**, 385 - 393. The crystal structure of glucose dehydrogenase from *Thermoplasma acidophilum.*

Johnson, S.L. & Smith, K.W. (1976) *Biochemistry* **15**, 3, 553 - 539. The interaction of borate and sulfite with pyridine nucleotides.

Jolley, K.A., Rappaport, E., Hough, D.W., Danson, M.J., Woods, W.G., Dyall-Smith, M.L., (1996) *J. Bacteriol.* **178**, 3044 - 3048. Dihydrolipoamide dehydrogenase from the halophilic archaeon *Haloferax volcanii*: homologous overexpression of the cloned gene.

Jolley, K.A., Russel, R.J.M., Hough, D.W., & Danson, M.J. (1997) *Eur. J. Biochem.* **248**, 362 - 368. Site-directed mutagenesis and halophilicity of dihydrolipoamide dehydrogenase from the halophilic archaeon, *Haloferax volcanii.*

Jones, T.A., Zou, J.Y., Cowan, S.W., & Kjeldgaard, M. (1991) *Acta Cryst.* **A47**, 110 - 119. Improved methods for building protein models in electron density maps and the location of errors in these models.

Jörnvall, H. Persson, B., & Jeffery, J. (1987) *Eur. J. Biochem.* **167**, 195 - 201. Characteristics of alcohol/polyol dehydrogenases.

Jörnvall, H., Von Bahr-Lindström, H., Jany, K-D., Ulmer, W., Fröschler, M. (1984) *FEBS Lett.* **165**, 190 - 196. Extended superfamily of short alcohol-polyol-sugar dehydrogenases: structural similarities between glucose and ribitol dehydrogenase.

K

Kandler, O. & König, H. (1985) *The Bacteria. A Treatsie on structure and function* **VII**, 413 - 457. Eds. Woese, C.R. & Wolfe, R.S.. New York - London, Academic Press. Cell envelopes of archaebacteria.

Karlsson, C., Maret, W., Auld, D.S., Höög, J-O., & Jörnvall, H. (1989) *Eur. J. Biochem.* **186**, 543 - 550. Variability within mammalian sorbitol dehydrogenases. The primary structure of the human liver enzyme.

Kelly, C.A., Nishiyama, M., Ohnishi, Y., Beppu, T., & Birktoft, J.J. (1993) *Biochemistry* **32**, 3912 – 3922. Determinants of Protein Thermostability Observed in the 1.9-Å Crystal Structure of Malate Dehydrogenase from the Thermophilic Bacterium *Thermus flavus.*

Kim, I., Ondrey, G., & Kamyia, T. (1998) *Chemical Engineering* **July**, 43 - 47. Betting Big on Biopolymers.

Kimblin, C., Hascall, T., & Parkin, G. (1997) *Inorg. Chem.* **36**, 5680 - 5681. Modelling the catalyic site of liver alcohol dehydrogenase: synthesis and structural characterization of a [Bis(thioimidazolyl)(pyrazolyl)hydroborato]zinc complex, $[HB(tim^{Me})_2pz]ZnI$.

Kimura, M., Arndt, E., Hatakeyama, T., Hatakeyama, T., & Kimura, J. (1989) *Can. J. Microbiol.* **35**, 195 - 199. Ribosomal proteins in bacteria.

Kirino, H., Aoki, M. Aoshima, M., Hayashi, Y., Ohba, M., Yamagishi, A., Wakagi, T., & Oshima, T. (1994) *Eur. J. Biochem.* **220**, 275 – 281. Hydrophobic interaction at the subunit interface contribute to the thermostability of 3-isopropylmalate dehydrogenase from an extreme thermophile, *Thermus thermophilus.*

Klingeberg, M., Galinsky, B., Sjoholm, C., Kaskhe, V., Antranikan, V. (1995) *Appl. Environ. Microbiol.* **61**, 3098 - 3104. Purification and properties of a highly thermostable, sodium dodecyl sulfate resistant, and stereospecific proteinase from the thermophilic Archaeon *Thermococcus stetteri.*

Knapp, S., De Vos, W., Rice, D., & Ladenstein, R. (1997) *J. Mol. Biol.* **267**, 916 - 932. Crystal structure of glutamate dehydrogenase from hyperthermophilic eubacterium *Thermotoga maritima* at 3.0 Å resolution.

Koch, R., Spreinat, A. Lemke, K., & Antranikian, G. (1991) *Arch. Microbiol.* **155**, 572 – 578. Purification and properties of a hyperthermoactive α-amylase from the archaeobacterium *Pyrococcus woesei.*

Kohen, A., Cannio, R., Bartolucci, S., & Klinman, J. (1999) *Nature* **399**, 496 - 499. Enzyme dynamics and hydrogen tunnelling in a thermophilic alcohol dehydrogenase.

König, H. & Stetter, K.O. (1986) *System. Appl. Microbiol.* **7**, 300 - 309. Studies on archaebacterial S-layers.

Koonin, E.V., Mushegian, A.R., Galperin, M.Y., & Walker, D.R. (1997) *Mol. Microbiol.* **25**, 4, 619 - 637. Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea.

Kraulis, P.J. (1991) *J. Appl. Crystallog.* **24**, 946 - 950. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures.

Kurz, L.C., Roble, J.H, Nakra, T., Drysdal, G.R., Buzan, J.M., Scwartz, B., & Drueckhammer, D.G. (1997) *Biochemistry* **36**, 3981 – 3990. Ability of Single-Site Mutants of Citrate Synthase To Catalyze Proton Transfer from the Methyl Group of Dethiaacetyl-Coenzyme A, a Non-Thioester Substrate Analog.

L

Laemmli, U.K. (1970) *Nature* **227**, 680 - 685. Clevage of structural proteins during the assembly of the head of bacteriophage T4.

Lake, J.A., Henderson, E., Oakes, M., & Clark, M.W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3786 - 3790. Eocytes: A new ribosome structure indicates a kingdom with a close relationship to eukaryotes.

Lake, J.A. (1988) *Nature* **331**, 184 - 186. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences.

Lake, J.A. & Rivera, M.C. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 2880 - 2881. Was the nucleus the first endosymbiont?

Lake, J.A. Jain, R. & Rivera, M.C. (1999) *Science* **283**, 2027 - 2028. Mix and match in the tree of life.

Lampel, K.A., Uratani, B., Chaudhruy, G.R., Ramaley, R.F., & Rudikoff, S. (1986) *J. Bacteriol.* **166**, 238 - 243. Characterization of the developmentally regulated *Bacillus subtilis* glucose dehydrogenase gene.

Lange, C.C., Wackett, L.P., Minton, K.W., & Daly, M.J. (1998) *Nature Biotech.* **16**, 929 – 933. Engineering a recombinant *Deinococcus radiodurans* for organopollutant degradation in radioactive mixed waste environments.

Langer, D., Hain, J., Thuriaux, P., & Zillig, W. (1995) *Proc. Natl. Acad. Sci. USA.* **92**, 5768 - 5772. Transcription in Archaea: similarity to that in Eukarya.

Langworthy, T.A. (1985) *in The Bacteria* **VIII**, 459 - 491. Eds. Woese, C.R. & Wolfe, R.S.. New York - London, Academic Press. Lipids of archaebacteria.

Lawler, A. (1995) *Science* **267**, 6, Walker bill to boost hydrogen sparks democratic grumbling.

Lebbink, J.H.G., Eggen, R.I.L., Geerling, A.C.M., Consalvi, V., Chiaraluce, R., Scandurra, R., & DeVos, W.M. (1995) *Prot. Eng.* **8**, 1287 - 1994. Exchange of domains of glutamate dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus* and the mesophilic bacterium *Clostridium difficile* : effects on catalysis, thermoactivity and stability.

Lee, S.Y. (1996) *Trends Biotechnol.* **14**, 431 - 438. Plastic bacteria? Progress and prospects for polyhydroxyalkanoate production in bacteria.

Lee, C.K. & Dobrogosz, W.J. (1965) *J. Bacteriol.* **90**, 3, 653 – 660. Oxidative metabolism in *Pedicoccus pentosaceus*.

Lepkowski, W. (1995) *Chem. Eng. News* **May**, 26, The politics of hydrogen.

Leslie, J. (1997) *Wired* **Oct.**, 1 -2. Dawn of the hydrogen age. (see appendices)

Lim, J-H., Yu, Y.G., Han, Y.S., Cho. S-J., Ahn, B-Y., Kim, S-H., & Cho, Y. (1997) *J. Mol. Biol.* **270**, 259 - 274. The crystal structure of an Fe-Superoxide Dismutase from the hyperthermophile *Aquifex pyrophilus* at 1.9 Å resolution: Structural basis for thermostability.

Little, J.W., Edmiston, S.H., Pacelli, L., & Mount, D.W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6, 3225 - 3229. Cleavage of the *Escherichia coli* lexA protein by the recA protease.

Little, J.W. & Mount, D.W. (1982) *Cell* **29**, 11 - 22. The SOS Regulatory System of *Escherichia coli*.

López-García, P. & Moreira, D. (1999) *Trends Biochem. Sci.* **24**, 88 - 93. Metabolic symbiosis at the origin of eukaryotes.

Lowry, O.H., Passonneau, J.V., & Rock, M.K. (1961) *J. Biol. Chem.* **236**, 10, 2756 - 2759. The stability of pyridine nucleotides.

Lübben, M. & Schäfer, G. (1989) *J. Bacteriol.* **171**, 11, 6106 - 6116. Chemiosmotic energy conversion of the archaebacterial thermoacidophile *Sulfolobus acidocaldarius*: oxidative phosphorylation and the presence of an Fo-related N,N'-dicyclohexylcarbodiiminde-binding proteolipid.

Luthi, E., Jasmat, N.B., & Bergquist, P.L. (1990) *Appl. Microbiol. Biotechnol.* **19**, 2677 - 2683. Overproduction of an acetyxylan esterase from the extreme thermophilic bacterium *Caldocellum saccharolyticum.*

Luthy, R., Bowie, J.U., & Eisenburg, D. (1992) *Nature* **356**, 83 - 85. Assessment of protein models with three-dimensional profiles.

M

Ma, K., Schicho, R.N., Kelley, R.M. & Adams, M.W.W. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5341 - 5344. Hydrogenase of the hyperthermophile *Pyrococcus furiosus* is an elemental sulfur reductase or sulfhydrogenase: Evidence for a sulfur-reducing hydrogenase ancestor.

Ma, K., Zhou, Z. H., & Adams, M.W.W. (1994) *FEMS Microbiol. Lett.* **122**, 245 - 250. Hydrogen production from pyruvate by enzymes purified from the hyperthermophilic archaeon, *Pyrococcus furiosus*: A key role for NADPH.

Mackness, R. (1991) Project *Report*, University of Bath.

Maloney, S.E., Marks, T.S., & Sharp, R.J. (1997a) *J. Chem. Technol. Biotechnol.* **68**, 357 - 360. Detoxification of synthetic pyrethroid insecticides by thermophilic microorganisms.

Maloney, S.E., Marsks, T.S., & Sharp, R.J. (1997b) *Lett. Appl. Microbiol.* **24**, 441 - 444. Degradation of 3-chlorobenzoate by thermophilic microorganisms.

Margulis, L. (1993) *Symbiosis in Cell Evolution*, W.H. Freeman.

Martin, W. & Müller, M. (1998) *Nature* **392**, 37 - 41. The hydrogen hypothesis for the first eukaryote.

Martins, L.O. & Santos, H. (1995) *Appl. Env. Microbiol.* **61**, 9, 3299 - 3303. Accumulation of mannosylglycerate and di-*myo*-inositol-phosphate by *Pyrococcus furiosus* in response to salinity and temperature.

Matthews, B. (1968) *J. Mol. Biol.* **33**, 491 - 497. Solvent content of protein crystals.

Matthews, B. (1993) Ann. Rev. *Biochem.* **62**, 139 - 160. Structural and genetic analysis of protein stability,

Matthews, R.G. & Massey, V. (1969) *J. Biol. Chem.* **244**, 7, 1779 - 1786. Isolation of Old yellow enzyme in free and complexed forms.

Matthews, B.W., Nicholson, H., & Becktel, W.J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6663 – 6667. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding.

Maynard Smith, J., Smith, N.H., O'Rourke, M., & Spratt, B.G. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4384 - 4388. How clonal are bacteria?

McCoy, M. (1999) *Chem. Eng. News* **January**, 10 – 14. Biocatalysis grows for drug synthesis.

McCutchen, C.M., Duffaud, G.D., Leduc, P., Petersen, A., Tayal, A., Khan, S.A., & Kelley, R.M. (1995) *Biotechnol. Bioeng.***52**, Purification, biochemical characterization and use of β-1,4-mannanase and A-1,6-galactosidase from the hyperthermophilic eubacterium *Thermotoga neopolitana*.

McPherson, A. (1990) *Eur. J. Biochem.* **189**, 1 - 23. Current approaches to macromolecular crystallization.

Menédez-Arias, L. & Argos, P. (1989) *J. Mol. Biol.* **206**, 397 – 406. Engineering Protein Thermal Stability. Sequence Statistics Point to Residue Substitutions in α-Helices.

Metzger, R.P., Wilcox, S.S., & Wick, A.N. (1964) *J. Biol. Chem.* **239**, 6, 1769 - 1772. Studies with Rat Liver Glucose Dehydrogenase.

Moore, J.C., Jin, H-C., Kuchner, O., & Arnold, F.H. (1997) *J. Mol. Biol.* **272**, 336 - 347. Strategies for the *in vitro* evolution of protein function: enzyme evolution by random recombination of improved sequences.

Morris, A.L., MacArthur, M.W., & Thornton, J.N. (1992) *Proteins* **12**, 345 - 364. Stereochemical quality of protein structure coordinates.

Mostad, S.B., Helming, H.L., Groom, C., & Glasfeld, A. (1997) *Biochem. Biophys. Res Comm.* **233**, 681 - 686. The stereospecificity of hydrogen transfer to $NAD(P)^+$ catalyzed by lactol dehydrogenases.

Mount, D.W., Walker, A. C., & Kosel, C. (1975) *J. Bacteriology* **121**, 3, 1203 -1207. Effect of *tsl* Mutations in Decreasing Radiation Sensitivity of a *recA*$^-$ Strain of *Escherichia coil* K-12.

Mevarech, M., Leicht, W., & Werber, M.M. (1976) *Biochemistry* **15**, 2383 - 2387. Hydrophobic chromatography and fractionation of enzymes from extremely halophilic bacteria using decreasing concentration gradients of ammonium sulfate.

N

Nevill-Manning, C.G., Wu, T.D., & Brutlag, D.L. (1998) *Proc. Acad. Sci. USA* **95**, 5865 - 5871. Highly specific protein sequence motifs for genome analysis.

Niehaus, F., Bertoldo, C., Kähler, M., & Antranikian, G. (1999) *Appl. Microbiol. Biotechnol.* **51**, 711 - 729. Extremophiles as a source of novel enzymes for industrial application.

O

Obon, J.M., Casanova, P., Manjon, A., Fernandez, V.M., & Iborra, J.L. (1997) *Biotech. Progress* **13**, 5, 557 - 561. Stabilization of glucose dehydrogenase with polythyleneimine in an electrochemical reactor with NAD(P)(+) regeneration.

Ó Fágáin, C. (1995) *Biochim. Biophys. Acta* **1252**, 1 - 14. Understanding and increasing protein stability.

Ohlsson, I., Nordström, B., & Bränden, C-I. (1974) *J. Mol. Biol.* **89**, 339 - 354. Structural and functional similarities within the coenzyme binding domains of dehydrogenases.

Oppenheimer, N.J., & Kaplan, N.O., (1974) *Biochemistry* **18**, 4675 - 4685. Structure of the primary acid rearrangement product of reduced nicotinamide adenine dinucleotide (NADH).

P

Pace, N.R (1992) *J. Mol.Biol.* **226**, 29 – 35. Contribution of the Hydrophobic Effect to Globular Protein Stability.

Pace, N.R., Olsen, G.J., & Woese, C.R. (1986) *Cell* **45**, 325 - 326. Ribosomal RNA phylogeny and the primary lines of evolutionary descent.

Pauly, H.E. & Pfleiderer, G. (1977) *Biochemistry* **16**, 4599 - 4604. Conformational and functional aspects of the reversible dissociation and denaturation of glucose dehydrogenase.

Pearson, W.R. & Lipman, D.J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444 - 2448. Improved tools for biological sequence comparison.

Pererseri, H., Holder, S., Sutherlin, D.P., Schwitter, U. King, D.S. & Schultz, P.G. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 258 - 261. A method for directed evolution and functional cloning of enzymes.

Pedroni, P., Mura, G.M. Galli, G., Pratesi, C., Serbolisca, L., & Grandi, G. (1996) *Int. J. Hydrogen Energy* **21**, 10, 853 - 858. The Hydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*: From basic research to possible future applications.

Persidis, A. (1998) *Nature Biotech.* **16**, 593 - 594. Extremophiles.

Persson, B., Krook, M., & Jörnvall, H. (1991) *Eur. J. Biochem.* **200**, 537 - 543. Characteristics of short chain alcohol dehydrogenases and related enzymes.

Persson, B., Zigler, Jr., J.S., & Jörnvall, H. (1991) *Eur. J. Biochem.* **226**, 15 - 22. A super-family of medium-chain dehydrogenases/reductases (MDR) Sub-lines including ζ-crystallin, alcohol and polyol dehydrogenases, quinone oxidoreductases, enoyl reductases, VAT-1 and other proteins.

Perutz, M.F. (1974) *Nature* **247**, 341 – 344. Stereochemical basis of heat stability in bacterial ferredoxins and in haemaglobin A2.

Perutz, M.F. (1978) *Science* **210**, 1187 - 1191. Electrostatic effects in proteins.

Perutz, M.F. & Raidt, H. (1975) *Nature* **255**, 256 – 259. Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2.

Potera, C. (1998) *Genet. Eng. News* **Feb**, 16 & 31. Extremophiles find a way into industry.

Pribnow, D. (1975) *J. Mol. Biol.* **99**, 419 - 443. Bacteriophage T7 Early Promoters: Nucleotide Sequences of Two RNA Polymerase Binding Sites.

Purcarea, C., Hervé, G., Ladjimi, M.M., & Cunin, R. (1997) *J. Bacteriol.* **179**, 13, 4143 - 4157. Aspartate transcarbamylase from the deep-sea hyperthermophilic Archaeon *Pyroccocus abyssi*: Genetic organization, structure, and expression in *Escherichia coli.*

Q

Querol, E., Perez-Pons, J.A., & Mozo-Villarias, A. (1996) *Prot. Eng.* **9**, 3, 265 – 271. Analysis of protein conformational characteristics related to thermostability.

R

Rao, S.T. & Rossman, M.G. (1973) *J. Mol. Biol.* **76**, 241 - 256. Comparison of super-secondary structures in proteins.

Radding, C.M. (1981) *Cell* **25**, 3 - 4. Recombination Activities of *E. coli* RecA Protein.

Raia, C.A., D'Auria, S., Guagliardi, A., Bartolucci, S., De Rosa, M., & Rossi, M. (1995) *Biosen. Bioelec.* **10**, 135 - 140. Characterization of redox proteins from extreme thermophilic archaebacteria: studies on alcohol dehydrogenase and thioredoxins.

Ramaswamy, S., Eklund, H., & Plapp, B.V. (1994) *Biochem.* **33**, 5230 - 5237. Structures of Horse liver alcohol dehydrogenase complexed with NAD^+ and substituted benzyl alcohols.

Reiter, W-D., Hüdepohl, U., & Zillig, W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9509 - 9513. Mutational analysis of an archaebacterial promoter: Essential role of a TATA box for transcription efficiency and start-site selection *in vitro.*

Rella, R., Raia, C.A., Pensa, M., Pisani, F.M. Gambacorta, A. De Rosa, M., & Rossi, M. (1987) *Eur. J. Biochem.* **167**, 475 - 479. A novel archaebacterial NAD^+-dependent alcohol dehydrogenase. Purification properties.

Remaut, E., Stanssens, P., & Friers, W. (1981) *Gene* **15**, 81 - 93. Plasmid vector for high efficiency expression controlled by the p_L promoter of coliphage lambda.

Robb, F.T., Park, J-B., & Adams, M.W.W. (1992) *Biochim. Biophys. Acta* **1120**, 267 - 272. Characterization of an extremely thermostable glutamate dehydrogenase: a key enzyme in the primary metabolism of the hyperthermophilic archaebacterium, *Pyrococcus furiosus.*

Roberts, T.M., Kacich, R., & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2, 760 - 764. A general method for maximizing the expression of a cloned gene.

Robinson, R. (1996) Project *Report*, University of Bath.

Rosenburg, A.H., Lade, B.N., Chui, D., Lin, S., Dunn, J.J., & Studier, F.W. (1987) *Gene* **56**, 125 - 135. Vectors for selective expression of cloned DNAs by T7 RNA ploymerase.

Rosazza, J.P.N., (1995) *ASM News* **61**, 5, 241 - 245. Biocatalysis, Microbiology, and Chemistry: The Power of Positive Linking.

Rossjohn, J.. (1994) *Ph. D. Thesis.* University of Bath.

Rossmann, M.G., Moras, D., & Olsen, K.W. (1974) *Nature* **250**, 194 - 199. Chemical and biological evolution of a nucleotide-binding protein.

Rost B. & Sander, C., (1993) *J. Mol. Biol.* **232**, 584 - 599. Prediction of protein structure at better than 70% accuracy.

Rost, B., Sander, C., & Schneider, R. (1994) *CABIOS* **10**, 53 - 60. PHD - an automatic mail server for protein secondary structure prediction.

Rost B. & Sander, C., (1994) *Proteins* **19**, 55 - 72. Combining evolutionary information and neural networks to predict protein secondary structure.

Rover, Jr., L., Fernandes, J.C.B., de Olivera Neto, G., Kubota, L.T., Katekawa, E., & Serrano, S.H.P. (1998) *Anal. Biochem.* **260**, 50 - 55. Study of NADH stability using ultraviolet-visible spectrophotometric analysis and factorial design.

Russel, R.J.M., Hough, D.W., Danson, M.J., Taylor, G.L. (1994) *Structure* **2**, 12, 1157 – 1167. The crystal structure of citrate synthase from the thermophilic Archaeon, *Thermoplasma acidophilum*.

Russel, R.J.M., Ferguson, J. M.C., Hough, D.W., Danson, M.J., & Taylor, G.L. (1997) *Biochemistry* **36**, 9983 – 9994. The Crystal Structure of Citrate Synthase from the Hyperthermophilic Archaeon *Pyrococcus furiosus* at 1.9 Å Resolution.

Russel, R.J.M., Gerike, U., Danson, M.J., Hough, D.W., & Taylor, G.L. (1998) *Structure* **6**, 3, 351 – 361. Structural adaptations of the cold-active citrate synthase from an Antartic bacterium.

S

Sambrook, J., Fritsch, E.F., & Maniatis, T. (eds) (1989) *Molecular cloning: a laboratory manual*, 2nd edn. Cold Spring Harbour Laboratory Press, New York.

Sanger, F., Nicklen, S., & Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 12, 5463 - 5467. DNA sequencing with chain-terminating inhibitors.

Schachter, H., Sarney, J., McGuire, E.J., & Roseman, S. (1969) *J. Biol. Chem.* **244**, 17, 4785 - 4792. Isolation of diphosphopyridine nucleotide-dependent L-fucose dehydrogenase from pork liver.

Schäfer, T, & Schönheit, P. (1992) *Arch. Microbiol.* **158**, 188 - 202. Maltose fermentation to acetate, CO_2 and H_2 in the anaerobic hyperthermophilic archaeon *Pyrococcus furiosus*: evidence for the operation of a novel sugar fermentation pathway.

Schäfer, G., Anemüller, S., Moll, R., Gleissner, M., & Schmidt, C.L. (1994) *Syst. Appl. Microbiol.* **16**, 544 - 555. Has *Sulfolobus* an archaic respiratory system? Structure, function and genes of its components.

Schäfer, G., Purchke, W.G., Gleissner, M., Schmidt, C.L. (1996) *Biochim. Biophys. Acta* **1275**, 16 – 20. Respiratory chains of archaea and extremophiles.

Schneider-Bernlöhr, H., Adolph, H-W., & Zeppenzauer, M. (1986) *J. Amer. Chem.* **108**, 5573 - 5576. Coenzyme specificity of alcohol/polyol dehydrogenases: conservation of protein types vs. functional constraints.

Scopes, R.K. (1983) *FEBS Lett.* **156**, 2, 303 - 306. An iron-activated alcohol dehydrogenase.

Scrutton, N.S., Berry, A., & Perham, R.N. (1990) *Nature* **343**, 38 - 43. Redesign of the coenzyme specificity of a dehydrogenase by protein engineering.

Segerer, A.H., Langworthy, T.A., & Stetter, K.O. (1988) *System. Appl. Microbiol.* **10**, 161 - 171. *Thermoplasma acidophilum* and *Thermoplasma volcanium* sp. nov. from Solfatara fields.

Segerer, A.H. & Stetter, K.O. (1991) *The Prokaryotes*, **1**, 2nd. edn., Springer Verlag, Ed. Belows, A. The Order Sulfolobales.

Sellek, G. & Chaudhuri, J.B. (1999) *Enzyme Microbiol. Technol.* **25**, 471 - 482. Biocatalysis in organic media by using enzymes from extremophiles.

Selig, M., Xavier, K.B., Santos, H., & Schönheit, P. (1997) *Arch. Microbiol.* **167**, 217 – 232. Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium *Thermotoga.*

Shanley, A. (1998) *Chem. Eng.* **July**, 63 - 66. Enzymes Usher in a New Era.

Shine, J. & Delgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4, 1342 - 1346. The 3'-Terminal Sequence of *Escherichia coli* 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites.

Shine, J. & Delgarno, L. (1975) *Nature* **254**, 34 - 37. Determinant of cistron specificity in bacterial ribosomes.

Shirakawa, M., Tsurimoto, T. & Matsubara, K. (1984) *Gene* **28**, 127 -132. Plasmid vectors designed for high-efficiency expression controlled by the portable *recA* promoter-operator of *Escherichia coli.*

Siebers, B., Wendisch, V.F., & Hensel, R. (1997) *Arch. Microbiol.* **168**, 120 - 127. Carbohydrate metabolism in *Thermoproteus tenax*: *in vivo* utilization of the non-phosphorylative Entner-Doudoroff pathway and characterization of its first enzyme, glucose dehydrogenase.

Smith, L.D., Budgen, N., Bungard, S.J., Danson, M.J., & Hough, D.W. (1989) *Biochem. J.* **261**, 973 - 977. Purification and characterization of glucose dehydrogenase from the thermoacidophilic archaebacterium *Thermoplasma acidophilum.*

Smith, L.D. (1989) *Ph. D. Thesis.* University of Bath.

Smith, M.W., Feng, D-F., & Doolittle, R.F. (1992) *Trends Biochem. Sci.* **17**, 489 -493. Evolution by acquisition: the case for horizontal gene transfers.

Sneddon, S.F. & Tobias, D.J. (1992) *Biochemistry* **31**, 2842 - 2846. The role of packing interactions in stabilizing folded proteins.

Sogin, M.L. (1991) *Curr. Opin. Genet. Dev.* **1**, 457 - 463. Early evolution and the origin of eukaryotes.

Somero, G. (1995) *Annu. Rev. Physiol.* **57**, 43 - 68. Proteins and temperature.

Sonawat, H.M., Srivastava, S., Swaminathan, S., & Govil, G. (1990) *Biochem. Biophys. Res Commun.* **173**, 358 - 362. Glycolysis and Entner-Doudoroff pathways in *Halobacterium halobium*: some new observations based on ^{13}C NMR spectroscopy.

Southern, E.M. (1975) *J. Mol. Biol.* **98**, 503 - 517. Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis.

Stemmer, W.P. (1994a) *Proc. Natl. Acad. Sci. USA* **91**, 10747 - 10751. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution.

Stemmer W.P. (1994b) *Nature* **370**, 389 - 391. Rapid evolution of a protein *in vitro* by DNA shuffling.

Stetter, K.O., Segerer, A., Zillig, W., Huber, G., Fiala, G., Huber, R., & König, H. (1986) *System. Appl. Microbiol.* **7**, 393 - 397. Extremely thermophilic sulfur-metabolizing Archaebacteria.

Stigter, D. & Dill, K. A. (1990) *Biochem.* **29**, 1262 - 1271. Charge effects on folded and unfolded proteins.

Stoichet, B.K., Baase, W.A., Kuroki, R., & Matthews, B.W. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 452 - 456. A relationship between protein stability and protein function.

Strecker, H.J. & Korkes, S. (1952) *Biol. Chem.* **196**, 769 - 784. Glucose Dehydrogenase.

Studier, F.W. & Moffatt, B.A. (1986) *J. Mol. Biol.* **189**, 113 - 130. Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes.

Studier, F.W., Rosenburg, A.H., Dunn, J.J., & Dubendorff, J.W. (1990) *Methods in Enzymology* **185**, 60 - 89.Use of T7 RNA Polymerase to Direct Expression of Cloned Genes.

Stura, E.A., & Wilson, I.A. (1992) *Crystallization of Nucleic acids and proteins, A practical approach*, Ed. Ducruix, A. & Giegé, R. **Chapter 5**, 99 - 126.

Sund, H. & Theorell, H. (1963) *The Enzymes* **7**, 25 - 83. Eds. Boyer, P.D., Lardy, H., & Myrbäck, K., 2nd edn. Academic Press, New York and London.

Sundquist, A.R. & Fahey, R.C. (1988) *J. Bacteriol.* **170**, 3459 - 3467. The novel disulfide reductase Bis-γ-glutamylcysteine reductase and dihydrolipoamide dehydrogenase from *Halobacterium halobium*: purification by immobilised metal ion affinity chromatography and properties of the enzymes.

T

Tanford, C. (1961) *Physical chemistry of macromolecules*, **Chapter 7**, Wiley, New York.

Tanford, C. (1968) *Adv. Protein Chem.* **23**, 121- 282. Protein denaturation.

Tanner, J.J., Hecht, R.M., & Krause, K.L. (1996) *Biochemistry* **35**, 2597 – 2609. Determinants of enzyme thermostability observed in the molecular structure of *Thermus aquaticus* D-glyceraldehyde-3-phosphate dehydrogenase at 2.5 Å Resolution.

Thaller, C., Eichele, G., Weaver, L.H., Wilson, E., Karlsson, R. & Jansonius, J.N. (1985) *Methods Enzymol.*, **114**, 132 - 135. Diffraction methods for biological macromolecules. Seed enlargement and repeated seeding.

Therorell, H. & Tatemoto, K. (1971) *Arch. Biochem. Biophys.* **142**, 69 - 82. Excitation transfer in complexes of horse liver alcohol dehydrogenase.

Trebbau de Acevedo, G. & McInerney, M.J. (1996) *J. Industr. Microbiol.* **16**, 1 - 7. Emulsifying activity in thermophilic and extremely thermophilic microorganisms.

Trent, S.D., Nimmesgern, E., Wall, J.S., Hartl, R.U., Horwich, A.L. (1991) *Nature* **354**, 490 - 493. A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein t-complex polypeptide-1.

Trent, S.D., Osipiuk, J., & Pinkau, T. (1990) *J. Bacteriol.* **172**, 1478 - 1484. Acquired thermotolerance and heat shock in the extremely thermophilic archaebacterium *Sulfolobus* sp. strain B12.

Tsai, C.J., Lin, S.L., Wolfson, UH.J., & Nussinov, R. (1997) *Protein Sci.* **6**, 53 - 64. Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect.

Tsurimoto,T., Hase, T., Matsubara, H., & Matsubara, K. (1982) *Mol. Gen. Genet.* **187**, 79 - 86. Bacteriophage lambda initiators: preparation from a strain that overproduces the O and P protein.

Turner, M.K. (1995) *Trends Biotechnol.* **13**, 173 - 177. Biocatalysis in organic chemistry (Part I): past and present.

Turner, M.K. (1995) *Trends Biotechnol.* **13**, 253 - 258. Biocatalysis in organic chemistry (Part II): present and future.

V

Van Den Burg, B., Vriend, G., Veltman, O.R., Venema, G., & Eijsink, V.G.H. (1998) *Proc. Natl. Acad. USA* **95**, 2056 – 2060. Engineering an enzyme to resist boiling.

Vielle, C., Hess, J.M., Kelly, R.M., & Zeikus, J.G. (1995) *Appl. Environ. Microbiol.* **61**, 1867 - 1875. xylA cloning and sequencing and biochemical characterization of xylose isomerase from Thermotoga neapolitana.

Vielle, C. & Zeikus, J.G. (1996) *Trends Biotechnol.* **14**, 183 – 190. Thermozymes: identifying molecular determinants of protein structural and functional stability.

Viikari, L., Kantelinen, A., Sundquist, J., & Linko, M. (1994) *FEMS Microbiol. News* **13**, 2-3, 335 - 350. Xylanases in bleaching - from an idea to the industry.

Vogt, G., Woell, S., & Argos, P. (1997) *J. Mol. Biol.* **269**, 631 - 643. Protein thermostability, hydrogen bonds and ion pairs.

W

Wallman, P.H., Thorsness, C.B., & Winter, J.D. (1998) *Energy* **23**, 4, 271 - 278. Hydrogen production from wastes.

Walsh, K.A.J., Daniel, R.M., & Morgan, H.W. (1983) *Biochem. J.* **209**, 427 - 433. A soluble NADH dehydrogenase (NADH: ferricyanide oxidoreductase) from *Thermus aquaticus* strain T351.

Wang, P.G., Fitz, W., & Wong, C-H. (1995) *Chemtech.* **April**, 22 – 32. Making complex carbohydrates via enzymatic routes.

Wenz, I., Loesche, W., Till, U., Petermann, H., & Horn, A. (1976) *J. Chromatogr.* **120**, 187 - 196. Purification and characterization of commercial NADH and accompanying dehydrogenase inhibitors.

Whittaker, R.H. (1969) *Science* **163**, 150 - 159. New concepts of kingdoms of organisms.

Wierenga, R.K., De Maeyer, M.C.H., & Hol, W.G.J (1985) *Biochemistry* **24**, 1346 - 1357. Interaction of pyrophosphate moieties with α-Helices in dinucleotide binding proteins.

Wierenga, R.K., Terpstra, P., & Hol, W.G.J. (1986) *J. Mol. Biol.* **187**, 101 - 107. Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint.

Wilson, J.G. & Newall, A.B. (1970) *General and inorganic chemistry*, second ed., Cambridge-Uni. Press, Cambridge.

Witkin, E.M. (1976) *Bacteriological Reviews* **40**, 4, 869 - 907. Ultraviolet Mutagenesis and Inducible DNA Repair in *Escherichia coli.*

Woodward, J., Mattingly, S.M., Danson, M., Hough, D., Ward, N., & Adams, M. (1996) *Nature Biotech.* **14**, 872 – 874. *In vitro* hydrogen production by glucose dehydrogenase and hydrogenase.

Woodward, J. & Orr, M. (1998) *Biotechnol. Prog.* **14**, 897 – 902. Enzymatic Conversion of Sucrose to Hydrogen.

Wong, C-H. & Whitesides, G.M. (1981) *J. Amer. Chem.Soc.* **103**, 4890 - 4899. Enzyme-catalyzed organic-synthesis - NAD(P)H cofactor regeneration by using glucose-6-phosphate and the glucose-6-phosphate dehydrogenase from *Leuconostoc mesenteroides.*

Woese, C.R & Fox, G.E. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 11, 5088 - 5090. Phylogenetic structure of the prokaryotic domain: the primary kingdoms.

Woese, C.R., Pace, N.P., & Olsen, G.J. (1986) *Nature* **320**, 401 - 402. Are arguments against archaebacteria valid?

Wu, J.T., Wu, L.H., & Knight, J.A. (1986) *Clin. Chem.* **32**, 2, 314 - 319. Stability of NADPH: effect of various factors on the kinetics of degradation.

Woese, C.R. & Fox, G.E. (1977) *Proc. Natl. Acad. Sci. USA* **74**,11, 5088 - 5090. Phylogenetic structure of the prokaryotic domain: the primary kingdoms.

Woese, C.R., Kandler, O., & Wheelis, M.L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4576 - 4579. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eukarya.

X

Xiao, L. & Honig, B. (1999) *J. Mol. Biol.* **289**, 1435 - 1444. Electrostatic contributions to the stability of hyperthermophilic proteins.

Y

Yip, K.S.P., Britton, L., Stillman, T.J., Lebbink, J., De Vos, W.M., Robb, F.T., Vetriani, C., Maeder, D., & Rice, D.W. (1998) *Eur. J. Biochem.* **255**, 336 – 346. Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the Glutamate Dehydrogenase family.

Yip, K.S.P., Stillman, T.J., Britton, K.L., Artymiuk, P.J., Baker, P.J., Sedelnikova, S.E., Engel, P.C., Pasquo, A., Chiraluce, R. & Consalvi, V. (1995) *Structure* **3**, 1147 - 1158. The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion pair networks in maintaining enzyme stability at extreme temperatures.

Z

Zatman, L.J., Kaplan, N.O., & Colowick, S.P. (1953) *J. Biol. Chem.* **200**, 127 - 212. Inhibition of spleen Diphosphopyridine nucleotidase by Nicotinamide, an exchange reaction.

Zillig, W., Schnabel, R., & Stetter, K.O. (1985) *Curr. Top. Microbiol. Immunol.* **114**, 1-18. Archaebacteria and the origin of the eukaryotic cytoplasm.

Zillig, W. (1991) *Curr. Opin. Genet. Dev.* **1**, 4, 544 - 551. Comparative biochemistry of Archaea and Bacteria.

Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., Lanzendörfer, M., & Kristjansson, J.B. (1994) *System. Appl. Microbiol.* **16**, 609 - 628. Screening for *Sulfolobales*, their plasmids and their viruses in Icelandic Solfataras.

Zwickl, P., Fabry, S., Bogedain, C., Haas, A., Hensel, R. (1990) *J. Bacteriol.* **172**, 4329 - 4338. Gylceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaebacterium *Pyrococcus woesei*: Characterization of the enzyme, cloning and sequencing of the gene, and expression in *Escherichia coli*.

Zuckerkandl, E. & Pauling, L. (1965) *J. Theor. Biol.* **8**, 2, 357 - 366. Molecules as documents of evolutionary history.

Appendices

1. Hydrogen & Fuel Cell Letter, (Feb, 1999) **XIV**, 2, 1 - 3. http://www.ttcorp.cpm/nha/thl/1999_feb.htm. Europe's first hydrogen gas station opens, renewable H_2 to come from Iceland.

2. Hydrogen & Fuel Cell Letter, (May, 1999) **XIV**, 2, 1 - 3. http://www.ttcorp.cpm/nha/thl/1999_ may.htm. Governor Davis launches California fuel cell partnership, dbb fuel cell engines starts new facility.

3. Hydrogen & Fuel Cell Letter, (Jun, 1999) **XIV**, 6, 1 - 3. http://www.ttcorp.cpm/nha/thl/1999_jun.htm. Five years in the making, US$18 Million hydrogen production/fuelling station opens in Munich.

A monthly newsletter on renewable fuels published by Peter Hoffmann.
Editor and Publisher: Peter Hoffmann Rhinecliff, NY 12574-0014.

Dawn of the Hydrogen Age: By Jacques Leslie

Cars that go 5,000 miles between fill-ups, electric power plants you buy like appliances, and a better standard of living. Automobile and power companies are spending billions to make it real.

I'm at the headquarters of Ballard Power Systems in Burnaby, a suburb of Vancouver, and my big fuel cell moment is about to occur. Following the example of the premier of British Columbia, the mayor of Chicago, and the chair of the Los Angeles Metropolitan Transit Authority, I am going to drink the exhaust from Ballard's prototype fuel cell municipal bus. This is less foolhardy than it sounds, since the only emission from the fuel cell engine is water. For this reason, many people think that fuel cells can change the world.

At Ballard, the exhaust-drinking routine has grown so tiresome that when I ask for a sip, Paul Lancaster, Ballard's treasurer, doesn't even offer me a glass: he suggests I cup my handsunder the bus's exhaust pipe. The pipe points straight down, presumably because its effluence is not a noxious gas that must be spewed into the atmosphere in hopes that it will dissipate. I bend over, and, within a few seconds, I collect several teaspoons of warm, clear liquid. As I begin to drink, I try to imagine a mountain stream, but the water is disappointingly bland. "Like distilled water," Lancaster explains, and I realize that what I'm drinking is, in a sense,

exactly that - the pure product of the union of hydrogen, the element that powers fuel cells, and oxygen in the engine. In some engine designs, even the exhaust water becomes an asset, recirculated to aid in internal processes. One tenet of the coming hydrogen age, according to businessman-cum-fuel cell visionary Joe Maceda, is that "pollution is a measure of inefficiency, and inefficiency is lost profit."

After decades of unfulfilled promise, fuel cell momentum is now so great that its emergence as a predominant technology appears just short of inevitable. During the early 1990s, nearly every major car manufacturer in the world launched a program to build a fuel cell automobile. Then, in April, a stunning announcement by Daimler-Benz AG suddenly gave the fuel cell age a timetable. Mercedes-Benz's parent company said it was investing US$145 million to buy a one-quarter interest in Ballard, the world's leader in fuel cell technology, and $150 million toward a joint venture with Ballard to create a new vehicle fuel cell engine company. Daimler-Benz also announced that beginning in 2005, the new company would produce 100,000 fuel cell engines annually. This is a remarkable figure, considering that the company, the world's 15th-largest auto manufacturer, makes only 700,000 cars a year now. "Daimler-Benz has a history of being one of the more conservative companies in the auto business, and arguably the auto business is one of the more conservative industries in the world," says Bill Reinert, a Toyota mechanical engineer. "So when somebody like Daimler pours millions of dollars into a technology and comes up making a statement like that, you have to say this might be pretty serious."

Though Daimler's fuel cell car will be powered by methanol, a hydrogen-rich derivative of natural gas, it is widely assumed that the use of fossil fuels to power fuel cells will be transitional, leading to an era in which hydrogen is extracted from sustainable energy sources. It is hard to overstate the implications of such a development: A drastic decline in air pollution, oil spills, acid rain, and greenhouse-gas emissions. An epochal geopolitical shift as global reliance on Middle East oil comes to an end and international trade balances are realigned.The emergence of quiet, decentralized electric plants sized according to need - small enough to power your car (and perhaps, at night, your house); big enough to power a town of 15,000 people, or, in tandem, a city. The disappearance of the electric grid is a possibility; a makeover of the electric-utility industry is nearly certain.

It is likely to take 50 to 100 years to achieve a mature "hydrogen economy," but the impact of fuel cells should be felt long before that. Over the next decade, products are likely to emerge on the market that are both more efficient and more environmentally benign than their predecessors. The century-long reign of internal-combustion engines will almost certainly be challenged by fuel cell-powered cars and buses that are quiet and clean, and use energy far more efficiently than today's vehicles. Naval forces in several countries are looking into fuel cells to run submarines and provide auxiliary power on oceangoing vessels; the US

Army is building a backpack-sizedfuel cell generator that can power a soldier's electronics gear, from night-vision goggles to infrared heat detectors. Fuel cell-driven desalination plants may offer clean water cheaply, defusing a potentially critical 21st-century resource shortage. Within a few years, fuel cells will probably power professional video cameras and many other products that now use batteries. Your laptop may eventually run on a fuel cell whose range is measured in days, not hours.